Universidade de Aveiro Departamento de Biologia
2014

Ana Rita Guimarães Rodrigues da Silva

Codon ambiguities as a mechanism to alter the genetic code in *Saccharomyces cerevisiae*

Ambiguidade de codões como mecanismo de alteração do código genético em *Saccharomyces cerevisiae*

Universidade de Aveiro Departamento de Biologia
2014

Ana Rita Guimarães Rodrigues da Silva

Codon ambiguities as a mechanism to alter the genetic code in *Saccharomyces cerevisiae*

Ambiguidade de codões como mecanismo de alteração do código genético em *Saccharomyces cerevisiae*

Tese apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Biologia Molecular e Celular, realizada sob a orientação científica da Doutora Ana Rita Macedo Bezerra, investigadora Pós-Doutoramento do RNA Biology Lab, Departamento de Biologia da Universidade de Aveiro

I would like to dedicate my thesis to my beloved grandfather. You gave me my tender memories and always had proud and seen the best of me. You will never be forgotten…

O júri

Presidente	Doutora Maria Adelaide de Pinho Almeida Professora auxiliar do Departamento de Biologia da Universidade do Aveiro
Arguente	Doutora Ana Catarina Gomes Diretora científica da Unidade de Genómica do Biocant
Orientadora científica	Doutora Ana Rita Macedo Bezerra Investigadora Pós-Doutoramento do RNA Biology Lab, Departamento de Biologia da Universidade de Aveiro

Agradecimentos

First, I would like to thank Professor Manuel Santos for giving me the opportunity to work in his laboratory and in this project. Also for receiving me so well and for the guidance and confidence in my work. A special thank you for broaden the horizons of my mind.

I am deeply grateful to my supervisor, Drª. Ana Rita Bezerra, for all the knowledge, advice, readiness, patience and specially, for the friendship built between us. Thank you for all the guidance and support during the course of this work. You are the best.

I would like to thank to all the extraordinary team from the RNA Biology Laboratory for welcoming me so well in the laboratory and for good working atmosphere. Thank you for all the support, fruitful discussions and for the good friendship. Also, I would like to thank to the past lab-member João Simões for the advice, helpful insights and support in the beginning of my project.

I would like to thank also to Dr. Yitzhak Pilpel for kindly providing the *Saccharomyces cerevisiae* deletion library used in this work. As well to University of Aveiro, specially the Biology Department, for providing me the conditions for the development of my Master Thesis.

Finally, but not least, a special thank you to my friends and my family, specially my mother, grandfathers and brother, for all the support and encouragement. Thanks to my friends from Aveiro for the love, support and joyful moments. A special thank you to my second family for their support, friendship and kind words.

Palavras-chave

Saccharomyces cerevisiae, código genético, tRNA, evolução, erros de tradução, ambiguidade do codão

Resumo

O código genético é geralmente visto como imutável, no entanto várias alterações à sua forma padrão são conhecidas. Uma das mais notáveis acontece em várias espécies do género *Candida*, onde o codão Leu-CUG é descodificado como serina por um novo RNA transferência (Ser-tRNA$_{CAG}$). O laboratório de acolhimento fez um grande progresso ao reverter a alteração atípica do código genético do fungo patogénico humano *C. albicans*, usando uma combinação de tRNAs mutantes, recombinação genética e evolução forçada. Estes resultados levantaram a hipótese que as ambiguidades sintéticas do codão, combinadas com evolução experimental, poderem libertar os codões do seu estado fixo.

Nesta tese testamos esta hipótese usando *S. cerevisiae* como modelo biológico. Geramos ambiguidade em codões específicos, de forma bifásica, envolvendo a deleção de genes de tRNA, seguida pela expressão de tRNAs não-cognatos capazes de compensar o tRNA eliminado. Tendo como base a ideia que os codões raros são mais suscetíveis a alterações que aqueles usados frequentemente, usamos duas estirpes *knock-out*, nas quais não existem os tRNAs cognatos capazes de descodificar os codões raros CUC-Leu e AGG-Arg.

Exploramos então a vulnerabilidade destes codões pela construção de tRNAs mutantes que incorporam erradamente Ser nestes locais. Estas estirpes recombinantes foram evoluídas ao longo do tempo, usando evolução experimental. Apesar de ter havido um forte impacto negativo na taxa de crescimento de estirpes que expressam o tRNA mutante a altos níveis, esta expressão a baixos níveis teve pouco impacto no *fitness* celular. Descobrimos que não só a ambiguidade do codão, mas também destabilizações da *pool* de tRNAs endógenos têm um impacto negativo na taxa de crescimento. Após a evolução, as estirpes com elevada expressão do tRNA mutante recuperaram significativamente em vários parâmetros de crescimento, o que mostra que estas adaptam-se e exibem maior tolerância à ambiguidade do codão. Através do sistema repórter fluorescente desenvolvido monitorizamos a incorporação errónea de Ser, o que nos indica que a Ser está de facto a ser incorporada e que, possivelmente, a alteração da identidade do codão foi atingida.

Apesar das consequências negativas gerais da ambiguidade do codão, demonstramos que os codões capazes de tolerar a perda do seu tRNA cognato, conseguem também tolerar a incorporação elevada de Ser. Isto levanta a hipótese que estes codões podem ser recodificados para outros aminoácidos naturais e/ou artificiais para a produção de proteínas com novas propriedades, contribuindo assim para o campo da Biologia Sintética e Biotecnologia.

Keywords

Saccharomyces cerevisiae, genetic code, tRNA, evolution, mistranslation, codon ambiguity

Abstract

Although the genetic code is generally viewed as immutable, alterations to its standard form occur in the three domains of life. A remarkable alteration to the standard genetic code occurs in many fungi of the Saccharomycotina CTG clade where the Leucine CUG codon has been reassigned to Serine by a novel transfer RNA (Ser-tRNA$_{CAG}$). The host laboratory made a major breakthrough by reversing this atypical genetic code alteration in the human pathogen *Candida albicans* using a combination of tRNA engineering, gene recombination and forced evolution. These results raised the hypothesis that synthetic codon ambiguities combined with experimental evolution may release codons from their frozen state.

In this thesis we tested this hypothesis using *S. cerevisiae* as a model system. We generated ambiguity at specific codons in a two-step approach, involving deletion of tRNA genes followed by expression of non-cognate tRNAs that are able to compensate the deleted tRNA. Driven by the notion that rare codons are more susceptible to reassignment than those that are frequently used, we used two deletion strains where there is no cognate tRNA to decode the rare CUC-Leu codon and AGG-Arg codon.

We exploited the vulnerability of the latter by engineering mutant tRNAs that misincorporate Ser at these sites. These recombinant strains were evolved over time using experimental evolution. Although there was a strong negative impact on the growth rate of strains expressing mutant tRNAs at high level, such expression at low level had little effect on cell fitness. We found that not only codon ambiguity, but also destabilization of the endogenous tRNA pool has a strong negative impact in growth rate. After evolution, strains expressing the mutant tRNA at high level recovered significantly in several growth parameters, showing that these strains adapt and exhibit higher tolerance to codon ambiguity. A fluorescent reporter system allowing the monitoring of Ser misincorporation showed that serine was indeed incorporated and possibly codon reassignment was achieved.

Beside the overall negative consequences of codon ambiguity, we demonstrated that codons that tolerate the loss of their cognate tRNA can also tolerate high Ser misincorporation. This raises the hypothesis that these codons can be reassigned to standard and eventually to new amino acids for the production of proteins with novel properties, contributing to the field of synthetic biology and biotechnology.

Contents

Chapter 1

1. Introduction 1

1.1 Genetic code 1

1.1.1 Reading the code 2

1.1.1.1 Transfer RNAs 2

1.1.1.2 Aminoacyl-tRNA synthetases 3

1.1.2 Protein biosynthesis 4

1.2 Alternative genetic codes 5

1.2.1 Evolution of the genetic code 7

1.2.2 Mechanisms of reassignment 9

1.2.3 Codon ambiguity and mistranslation 10

1.3 *Saccharomyces cerevisiae* as a biological model 12

1.4 Working hypothesis and objectives 13

Chapter 2

2. Materials and methods 15

2.1 Strains and growth conditions 15

2.1.1 Strains 15

2.1.2 Growth conditions 15

2.2 Oligonucleotides 16

2.3 Plasmids 17

2.3.1 Original plasmids 17

2.3.2 Constructed plasmids 17

2.3.3 Site directed mutagenesis 19

2.4 Preparation of *E. coli* competent cells 20

2.5 Plasmidic DNA purification from *E. coli* 20

2.5.1 Miniprep kit 20

2.5.2 “Homemade” minipreps 20

2.6 Plasmidic DNA purification from *S. cerevisiae* 21

2.7 Transformation procedures 21

2.7.1 *Escherichia coli* 21

2.7.2 *Saccharomyces cerevisiae* 21
2.7.3 *Saccharomyces cerevisiae* (alternative) 22
2.8 Colony PCR 22
2.9 Growth curves and rates 22
2.10 Evolution experiments 23
2.11 Northern Blot Analysis 23
2.11.1 Total RNA extraction 23
2.11.2 Northern blot 24
2.12 Epifluorescence microscopy 25

Chapter 3

3. Characterization of codon ambiguities 27
3.1 Overview 27
3.2 Results 29
3.2.1 Expression of mutant Ser-tRNAs in *S. cerevisiae* 29
3.2.2 Leucine set 31
3.2.3 Arginine set 32
3.2.4 Impact of CUC and AGG ambiguity 34
3.3 Discussion 37

Chapter 4

4. Evolution of codon ambiguities 41
4.1 Overview 41
4.2 Results 42
4.2.1 Evolution of the Leucine set 42
4.2.2 Evolution of the Arginine set 45
4.2.3 Northern blot analysis of mutant tRNA expression 47
4.2.4 Mutant tRNA sequencing 48
4.2.5 Monitoring of Ser misincorporation 50
4.2.6 Discussion 54

Chapter 5

5. Discussion 57

5.1 General discussion and future conclusions 57

5.2 Future work 59

List of abbreviations

5-FOA – 5-Fluoro-orotic Acid

aa-AMP – Aminoacyl-adenylate

aaRS – Aminoacyl-tRNA synthetase

aa-tRNA – Aminoacylated tRNA

AMP – Adenosine monophosphate

ATP – Adenosine triphosphate

BiP – Immunnglobulin-binding protein

eIF – Eukaryotic translation initiation factors

ER – Endoplasmatic reticulum

eRF – Eukaryotic release factors

f^5C – 5-formylcytidine

GFP – Green fluorescent protein

m^7G_{34} – 7-methylguanosine

MC groups – Strains constructed with multi-copy vectors

mRNA – Messenger RNA

OMP – Orotidine-5'-phosphate

PKA – cAMP-protein kinase A

Pyl – Pirrolysine

SC groups – Strains constructed with single-copy vectors

SCS – Stop codon suppression method

Sec – Selenocisteine

SPI – Supplementation based incorporation method

t^6A_{37} – N6-threonylcarbamoyladenosine

tRNA – Transfer RNA

WT – Wild-type

yEGFP – Yeast-enhanced green fluorescent protein

ΔtL – Abbreviation of the knock-out strain ΔtL(GAG)G

ΔtR – Abbreviation of the knock-out strain ΔtR(CCU)J

ψ – Pseudouridine

List of figures

Figure 1.1 Standard genetic code

Figure 1.2 tRNA structure and translation overview

Figure 1.3 mRNA translation

Figure 1.4 Genetic code diversity

Figure 1.5 Mechanisms of codon reassignments

Figure 1.6 Delivery of aminoacylated tRNAs translational machinery

Figure 3.1 Aminoacylation with canonical (a) and non-canonical amino acids for protein translation

Figure 3.2 Structure of serine, leucine and arginine

Figure 3.3 Representation of mutant tRNAs in the strain background of this work

Figure 3.4 Impact of mutant tRNAs on cell fitness of ΔtL(GAG) mutants

Figure 3.5 Impact of mutant tRNAs on cell fitness of ΔtR(CCU) mutants

Figure 3.6 Leu-CUC (A) and Arg-AGG (B) codon distribution over *S. cerevisiae* genome

Figure 3.7 CUC and AGG codon distribution over the PKA related genes group (A) and a schematic of the Ras/cAMP/PKA pathway (B)

Figure 4.1 Different types of experimental evolution

Figure 4.2 Comparison between the growth rate of non-evolved (grey) and evolved (black) Leucine mutants

Figure 4.3 Comparison between the lag phase duration, in hours, of non-evolved (dark grey) and evolved Leucine mutants (light grey)

Figure 4.4 Comparison between the growth rate of non-evolved ΔtL strains highly expressing the endogenous Ser-tRNA$_{UGA}$ (blue bar), and evolved strains with 100 (orange bar) and 200 generations (grey bar)

Figure 4.5 Comparison between the growth curve profiles of non-evolved (ΔtL, light blue) and evolved (EΔtL) Leu mutants

Figure 4.6 Comparison between the growth rate of non-evolved (grey) and evolved (black) Arginine (ΔtR) mutants

Figure 4.7 Comparison between the lag phase duration, in hours, of non-evolved (dark grey) and evolved (light grey) Arginine mutants

Figure 4.8 Northern blot analysis of the mutant tRNAs

Figure 4.9 Sequencing data from strains highly expressing Ser-tRNA$_{GAG}$ and the theoretical structure of the mutant tRNA

Figure 4.10 Scheme of the reporter system built to monitor serine incorporation at AGG-Arg codons

Figure 4.11 Scheme of the reporter system built to monitor serine incorporation at CUC-Leu codons

Figure 4.12 Quantification of the mean fluorescence of evolved Leu mutants expressing Ser-tRNA$_{GAG}$

Figure 4.13 Area measurements from non-evolved and evolved Leu mutants

Figure 4.14 Representative pictures of non-evolved and evolved Arg mutants

List of tables

Table 2.1 Yeast strains used in this study

Table 2.2 List of oligonucleotides used in this thesis

Table 2.3 Original plasmids used in this study

Table 2.4 List of the engineered plasmids

Table 2.5 List of oligonucleotides used for northern blot analysis

Table 3.1 Principal characteristics of amino acids involved in this study

Table 3.2 PKA related genes with 6 or more CUC codons

Chapter 1

1. Introduction

1.1 Genetic code

The genetic code was established 3.5 billion years ago (1) and is one of the oldest and most conserved characteristics of life. It is universally used across the three domains of life and can be defined as a series of biochemical reactions that establish the system of rules by which the information encoded by a nucleotide sequence is translated into a protein. Each three-nucleotide combination represents a codon which encodes a single amino acid. Since there are 4 nucleotides, there is a total of 64 combinations where 3 of them represent the termination codons UGA, UAA and UAG, and the remaining 61 encode for the 20 canonical amino acids (figure 1.1). This means that most amino acids are specified by more than one codon, showing the redundancy of the code. Codons that encode the same amino acid are named synonymous codons and are important to minimize the harmful effects of potential incorrectly placed nucleotides (2). However these synonymous codons are not equivalent and are used with different frequencies in high or low expressing genes, and in different organisms (3). This codon usage biases are linked with gene expression levels, translation efficiency and protein folding (4).

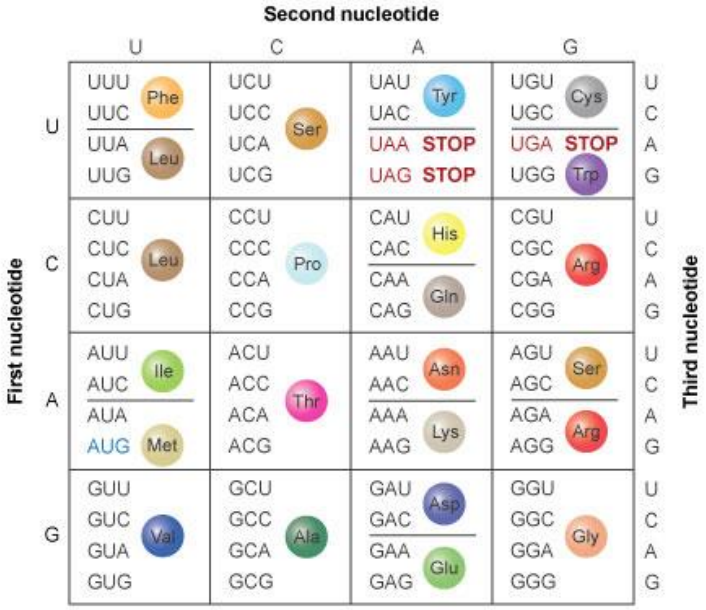


Figure 1.1 – Standard genetic code. The genetic code is composed of 64 codons and each one is composed of 3 nucleotides. The initiation codon is highlighted as blue (AUG) and termination codons are highlighted as red (UAA, UAG, UGA), and the rest of the codons show their specified amino acid. Adapted from Clancy and Brown, 2008.

1.1.1 Reading the code

Codon assignments are established by two major components: transfer RNAs (tRNA) and aminoacyl tRNA synthetases (aaRS). Codons are recognized by the complementary anticodon of its cognate tRNA, which in turn is aminoacylated by aaRS that are highly selective for the correspondent amino acid (6). This relationship is central to the genetic code and pivotal to protein biosynthesis (figure 1.2).

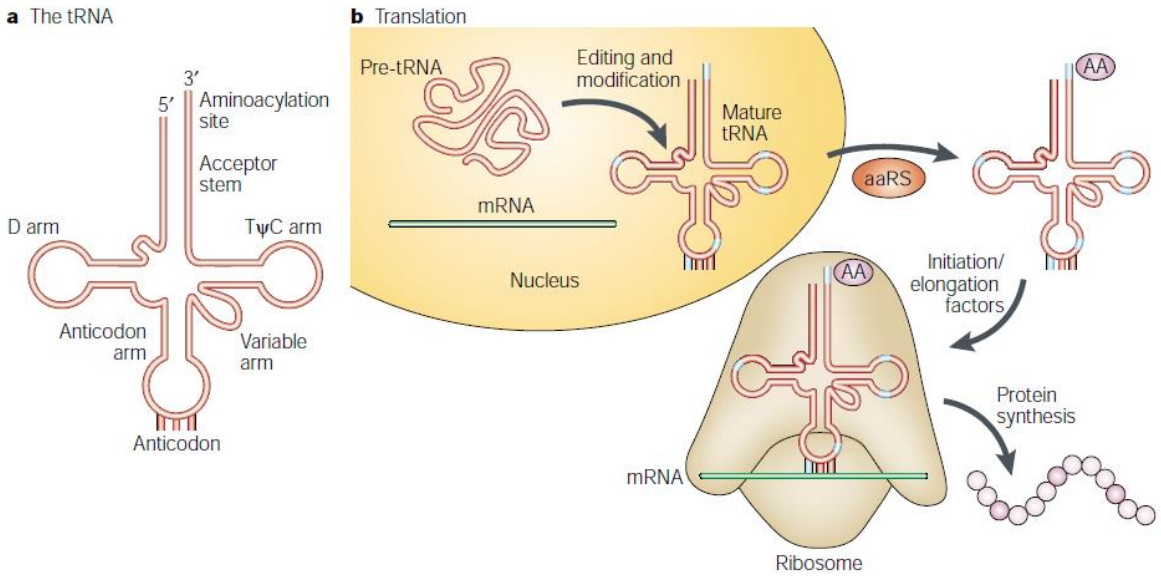


Figure 1.2 – tRNA cloverleaf structure (a) and translation overview (b) highlighting the role of tRNAs in translation, and aminoacyl-tRNA synthetases (aaRS) in their aminoacylation. Adapted from Knight *et al.*, 2001

1.1.1.1 Transfer RNAs

Transfer RNAs are the interface between the genetic information encoded in messenger RNA (mRNA) and proteins. tRNAs are small molecules composed of a single polynucleotide chain with 73-90 nucleotides which are arranged in a tridimensional L-shaped structure. Its secondary structure is often represented as a cloverleaf structure (figure 1.2a) composed by the acceptor stem, D arm, TψC arm, variable arm and the anticodon arm (7). The acceptor stem is composed of 7 base pairs followed by an unpaired nucleotide at position 73 and the 3'-CCA terminal where the amino acid is attached, whereas the TψC and anticodon arms have a 5 base pair stem and a 7 nucleotide loop. Stems are stabilized by Watson-Crick interactions and an occasional wobble G-U base pair is often observed. The major deviations seen in tRNA size are due to variations in nucleotide number in the variable and D loops. (8). One important structural motif is the anticodon

U-turn that involves the ubiquitous U_{33}. This is responsible for an abrupt reversion in direction (~180°) of the tRNA chain which exposes the anticodon nucleotides to anticodon-codon pairing during posterior mRNA decoding (9). In turn, so that the tRNA is aminoacylated, aaRS must recognize the correct tRNA which involves a series of identity elements.

Transfer RNAs have many post-transcriptional modifications and are by far the most and more diversely modified biological molecule. While some modifications simply involve the addition of a methyl group, others are rather complex and involve multi-step reactions catalyzed by a series of specialized enzymes. These ensure tRNA proper folding and function, and can be divided in two major groups: those that affect structure, and those that tackle critical positions to mRNA decoding and aaRS recognition (10). Although modified nucleotides can be present all over the tRNA molecule, the two most frequent modifications are found in positions 34 and 37 present in the anticodon loop. For example, modifications in position 37 aid in the maintenance of the U-turn and influence frameshifting, and therefore have an impact in translational accuracy (11). Modifications in the wobble position 34 can account for some of the degeneracy of the genetic code; for example an U_{34} enables stacking with A and G nucleotides (the latter by wobble interaction), while a modified I_{34} (inosine-34) enables recognition of C, A and U nucleotides, thus altering tRNA decoding capacity (12). Generally, modifications in these two positions contribute to efficient anticodon-codon pairing and to stability of the tRNA-mRNA interaction during mRNA translation (11). Also, some modified nucleotides in the anticodon domain are crucial to aaRS's ability to accurately aminoacylate tRNAs (13).

Noteworthy, tRNAs have other parts on the biological theatre beyond its role as adaptor in protein synthesis, not only in its aminoacylated state but also in its uncharged form, and even in fragmented form (7). Uncharged tRNAs have been shown to act as molecular sensors of external stresses such as amino acid deprivation, which regulates global gene expression to counteract nutritional stress (14, 15). Beside the canonical function of aminoacyl-tRNAs, they have also been implicated in other non-ribosomal functions. In bacterial cell wall formation, they act as substrates for building peptidoglycan bridges (16); and also mediate aminoacylation of phospholipids in the bacterial cell membrane (17). They also have a role in antibiotic biogenesis and in protein labeling for degradation (7).

1.1.1.2 Aminoacyl-tRNA synthetases

Aminoacyl-tRNA synthetases catalyze the aminoacylation reaction, which corresponds to the attachment of the correct amino acid to its cognate tRNA. This occurs through a two-step mechanism by first activating the amino acid and then transferring it to the tRNA. The aaRS binds and activates the amino acid with ATP, forming aminoacyl-adenylate (aa-AMP), and then this complex recognizes the cognate tRNA and transfers the amino acid to its 3' end, releasing the AMP. This is a rather complex task for aaRS, as they have to recognize the correct tRNA (between a large family of structurally similar tRNAs), as well as discriminate amino acids (18). Despite this, aaRS have an overall error rate of 10^{-4} which is achieved by a series of quality control mechanisms (19). There is one aaRS for each one of the 20 amino acids, and their recognition is usually done in two phases: the first activating step is inaccurate and is followed by a second step where non-cognate amino acids are hydrolyzed by an intrinsic proofreading activity, namely editing. While some amino acids have quite distinct physical and chemical properties which facilitates their discrimination, some amino acids, such as valine and isoleucine, only differ by a methyl group and their discrimination cannot be accomplished without the aaRS editing activity (20). On the other hand, tRNA is a large molecule that allows a series of intricate contacts with aaRS during recognition, mainly with the discriminator base (N73), the acceptor stem, and the anticodon. These so called identity elements can be determinants or antideterminants whether they promote or prevent, respectively, the interaction between the aaRS and the tRNA. By so, aminoacyl tRNA synthetases establish the genetic code by accurately pairing cognate tRNAs with their corresponding amino acids (19).

1.1.2 Protein biosynthesis

After the transcription of a particular gene, the synthesis of its protein product can begin (figure 1.3). Translation can generally be divided in three major steps: initiation, elongation and termination. This process is orchestrated by the ribosome, together with several auxiliary factors that bind to mRNA.

Initiation begins with the loading of the initiator methionyl-tRNA (Met-$tRNA_i$) in the ribosomal subunit 40S P site, forming a 43S pre-initiation complex. This is accomplished by eIF2·GTP complex, together with the eukaryotic translation initiation factors 1, 3 and 5 bound to the ribosomal subunit. With the help of eIF4F, the complex is positioned onto the 5' end of a capped and polyadenylated mRNA and scans it until an initiator codon AUG is found, which in turn, is recognized by the Met-

$tRNA_i$. Assemblage of the ribosomal subunit 60S to the later complex is facilitated by eIF3, eIF1 and 1A, forming the ribosome and enabling the elongation step. While Met-$tRNA_i$ is bounded to peptidyl (P) site, another aminoacylated tRNA is delivered by eEF1A·GTP to the acceptor (A) site of the ribosome. If the latter is the cognate tRNA, a peptide bond between the carried amino acid and the previous amino acid is formed. The deacylated tRNA is released and the complex is translocated, exposing the next codon to the A site and the process is repeated, codon by codon. When a termination codon is encountered, the eukaryotic release factor 1 (eRF1) binds to the A site, which triggers ribosome arrest and the finished peptide is released. With the aid of eRF3, the ribosomal subunits, along with deacylated tRNAs and auxiliary factors dissociate from mRNA, and are free to initiate another round of translation (21, 22).

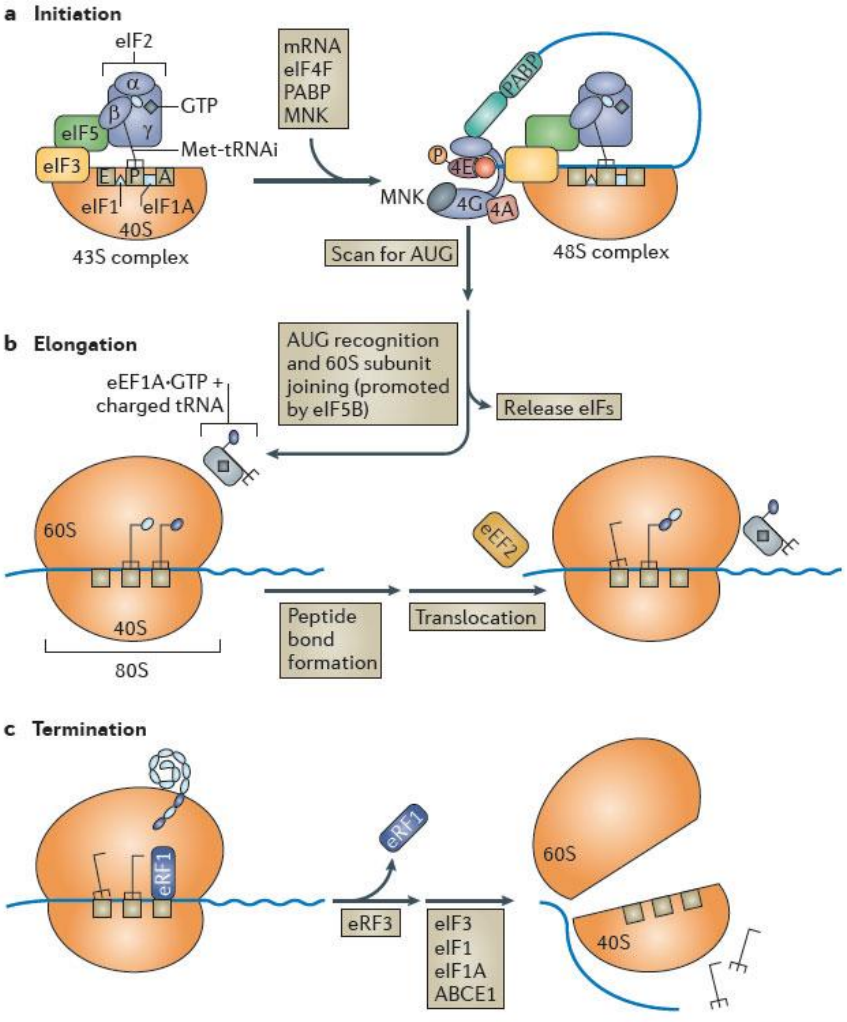


Figure 1.3 - mRNA translation. This process has three main phases: initiation (a), elongation (b) and termination (c), which requires a series of specific translation factors. Adapted from Walsh and Mohr, 2011

1.2 Alternative genetic codes

The genetic code is universally used by all forms of life and was initially postulated by Crick as a "frozen accident" unable to further evolve, since any alteration to the code would produce aberrant proteins, leading to proteome mayhem that would be lethal to the cell (2). Nevertheless, several natural deviations from its standard form are known in various microorganisms and mitochondria (figure 1.4). Together with the expansion of the amino acid repertoire by incorporation of two natural non-canonical amino acids (23), it shows that the code is not as frozen as initially thought.

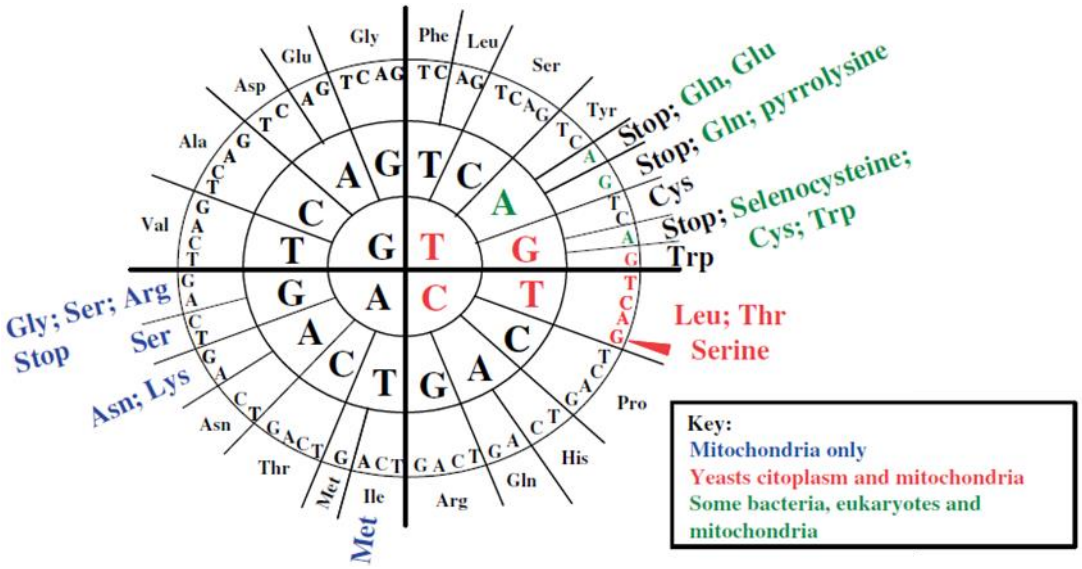


Figure 1.4 – Genetic code diversity found in mitochondria and in eubacteria, archaea and eukaryotic cytoplasm. Unchanged codons are in black. Adapted from Miranda *et al.*, 2006

The first natural alterations were discovered in human mitochondria and involved the decoding of the UGA stop codon as Trp, and the AUA-Ile codon as Met (24). Subsequent discoveries have shown that mitochondria are rather prone to codon identity reassignments and significant diversity of nonsense-to-sense and sense-to-sense reassignments is often found in Euglenozoans, Haptophytes, Stramenophiles, alveolates, green plants, red algae, fungi and metazoans (23). There are 16 known alterations in mitochondria and the rationalization is that they are particularly tolerant to reassignments due to their reduced genome size and complexity, when compared to nuclear genomes. Still, 10 alterations in nuclear codes were also found and termination codons are usually the only ones that are reassigned by the eukaryotic cytoplasmic translational machineries (24). For example, the UGA stop codon has been reassigned to Cys in *Euplotes* spp. (25) and to Trp in the *Colpoda* sp. and in various heterotrichs (26). The UUA and UAG stop codons have also been

reassigned to Gln in diplomonads (27), ciliates (26) and green algae (28). Termination codons appear to be particularly flexible, as they are also the target for the incorporation of the non-canonical amino acids selenocysteine, in a wide range of prokaryotes and eukaryotes (29) , and pyrrolysine in archaeal Methanosarcina species (30), producing novel classes of proteins with unique catalytic properties. The only known sense-to-sense reassignment in nuclear genomes is found in many fungi of the Saccharomycotina CTG with the decoding of the Leu-CUG codon as Ser (31, 32). Another interesting observation in nuclear genomes is that, in a small number of species, some codons like the Arg-AGA and Ile-AUA codons in *Micrococcus* spp., and the Arg-CGG codon is *Mycoplasma capricolum* are either extremely rare or absent, and appear unassigned (33). All together, these observations invalidate the hypothesis of a non-mutable code and shows that it is flexible and evolvable.

1.2.1 Evolution of the genetic code

Codon reassignments are an intriguing biological brainteaser since they show that the genetic code is still evolving, although these events are strongly selected against due to their intrinsic negative impact in proteome stability. Two evolutionary theories have arisen to explain how these alterations could emerge without leading to species extinction – the codon capture theory and the ambiguous intermediate theory (figure 1.5) (34).

The codon capture theory (figure 1.5.a) was initially proposed by Osawa and Jukes as a neutral mechanism in which genetic code changes arise from fluctuations in the GC/AT balance in the evolving genome. It postulates that evolution of codon reassignment happens through a stage in which a particular codon disappears from the genome, possibly leading to the disappearance of the cognate tRNA. At a later time in evolution, the codon may reappear by genetic drift and be decoded through a non-cognate tRNA, misreading the codon and thus allowing its identity alteration (24, 35). The disappearance of CGG-Arg codon from the AT-rich genome (75%) of *Mycoplasma capricolum*, rendered the hypothesis that an AT-rich genome has a genomic bias against GC content, which can drive GC-rich codons to disappear. This is further supported by the similar disappearance of AGA-Arg and AUA-Ile codons from the GC-rich genome (74%) of *Micrococcus luteus* (34). However, there are reassignments in genomes with no obvious GC or AT bias, and even cases where reassignments occurs against such bias, which questions the validity of this theory (36).

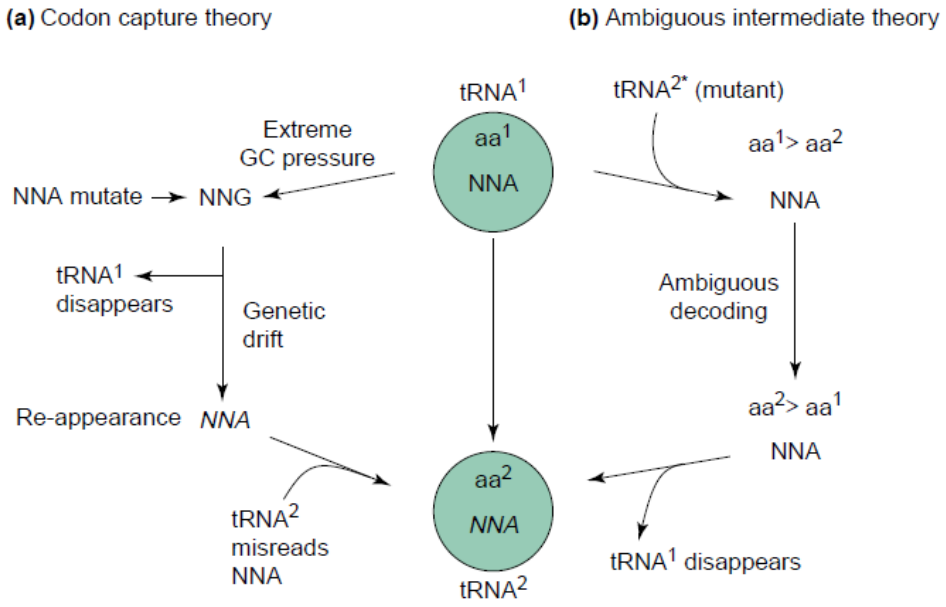


Figure 1.5 – Mechanisms of codon reassignments, where (a) illustrates the codon capture theory and (b) the ambiguous intermediate theory. Adapted from Santos *et al.*, 2004

Alternatively, Schultz and Yarus proposed the ambiguous intermediate theory (figure 1.5.b) as a non-neutral mechanism driven through tRNA mutations (37). It postulates that codon reassignments can appear through an intermediate stage of codon ambiguity when a single codon is both decoded by its cognate tRNA and a mutant tRNA. This translational ambiguity is the initial step for a gradual codon reassignment, and if selected, will lead to a gradual decrease of the cognate tRNA usage, and eventually to its loss. (34, 37). This theory applies both to sense and nonsense codon reassignments, but in the latter case the termination codon is being disputed by the release factor (RF) and a mutant tRNA capable of decoding it (37). Since codon ambiguity leads to proteome wide amino acid misincorporation and consequently to the synthesis of aberrant proteins, it is expected to be strongly selected against, as it would be highly detrimental to cell survival. However, this theory is credited by the ambiguous decoding of the CUG-Leu codon as Ser in several extant *Candida* species (32) and by the dual meaning of the *Bacillus subtilis* UGA codon as Trp or stop (38). Contrary to codon capture theory, this does not impose codon disappearance but it states that rarely used codons are more prone to reassignment than frequently used ones, so the potential negative impact of codon ambiguities in the proteome is minimized (37).

1.2.2 Mechanisms of reassignment

The molecular mechanisms of codon reassignment are poorly understood, however several studies provide interesting insight on how genetic code alterations may occur. They link them to alterations in components of the translational machinery accountable for interpreting the genetic code, namely tRNAs, aminoacyl-tRNA synthetases, and the release factors that recognize stop codons (39). As presented above (figure 1.4), some codons appear reassigned more frequently than others, particularly termination codons. Stop codon alterations are associated with critical structural changes in the codon recognition domain of release factors and in the anticodon of the misreading tRNAs. Such is the case of UAA and UAG decoding as Gln and UGA decoding as Trp or Cys, which occurs in several species of ciliates (39). It is proposed that this kind of reassignment happens through a first stage of codon ambiguity due to the appearance of a mutant tRNA capable of reading stop codon(s), and so competing with the RF. As ambiguity increases, the RF would fail in its ability to recognize the respective codon, enabling full reassignment (34). Another interesting way of exploiting the code was found by yeast mitochondria, where the four leucine CUN codons are decoded as threonine. This came as result of the loss of the Leu-$tRNA_{UAG}$ capable of decoding the CUN codons and the appearance of a mutant Thr-$tRNA_{UAG}$ with an atypically large anticodon loop. Surprisingly, this mutant Thr-tRNA has evolved from a His-tRNA, in which a single-nucleotide change converted it to a substrate for the yeast mitochondrial threonyl-tRNA synthetase (40). A distinct mechanism involves the reassignment of the nuclear Leu-CUG codon to Ser in various *Candida* species, where a mutation in the Ser-$tRNA_{CAG}$ produced a novel tRNA that is recognized both by SerRS and LeuRS. This renders the CUG codon ambiguous, as it is decoded both as Ser and Leu *in vivo* (41).

An important question raised by these studies is if, during the course of codon reassignment, tRNA selection can affect codon usage in a way that might lead to codon final reassignment. In this way, we have to consider two key aspects: tRNA abundance which is correlated with codon usage, and the strength of codon-anticodon interaction. Since codon reassignments tend to happen in less abundant codons, they are also translated by less abundant tRNAs (34) and such rarity renders them more vulnerable to competition during translation if a novel non-cognate tRNA would emerge (42). Strength of codon-anticodon interactions is modulated by tRNA structure and modified nucleosides, particularly those in the anticodon loop at positions N34, N37, and in some cases N35. Modified nucleosides in these positions have impact in codon recognition, and some have the potential to restrict the decoding capacity of tRNAs, while others expand it (39). For example, in squid mitochondria, Ser-$tRNA_{GCU}$ contains m^7G_{34} which expands its capacity to read Arg-AGA and

AGG codons, inserting serine at these sites (43). The mitochondrial Asn-tRNA of echinoderms has a 5'-Gψ_{35}U-3' anticodon which enables it to recognized AAA-Lys codons. Also, the decoding of mitochondrial AUA-Ile codons as methionine by Met-tRNA$_{CAU}$ is enabled by t^6A_{37} which stabilizes codon-anticodon interaction (34).

1.2.3 Codon ambiguity and mistranslation

Maintaining the integrity of the proteome is essential for cell viability as erroneous translation can induce protein misfolding and aggregation, and eventually cell death (figure 1.6). So codon ambiguities pose a striking problem to cellular homeostasis as they produce statistical proteins.

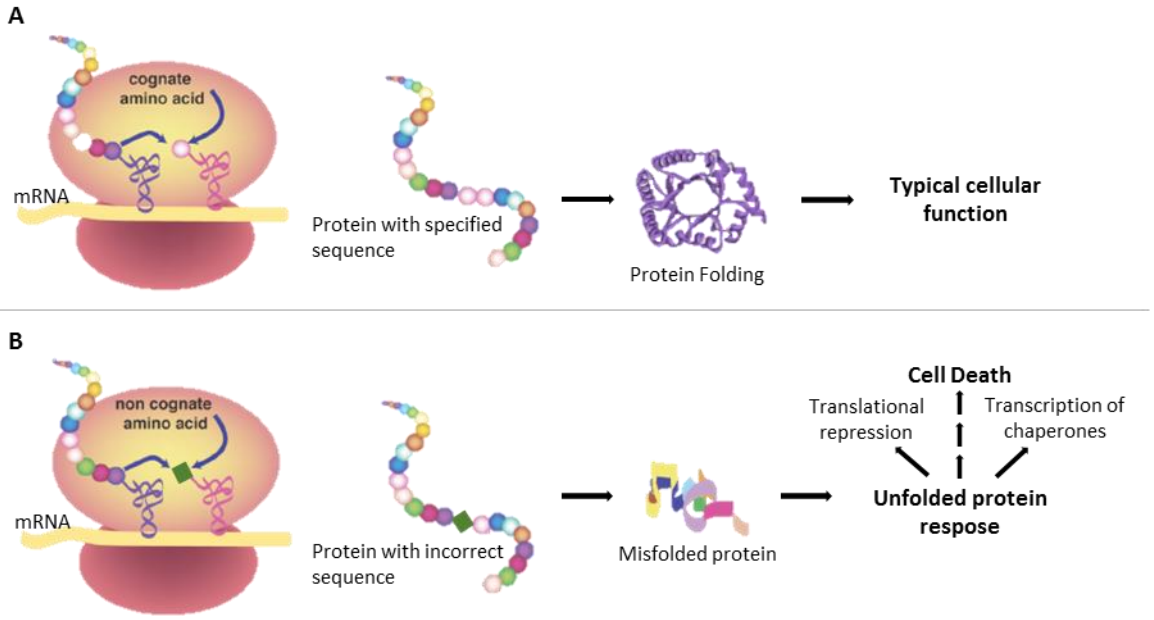


Figure 1.6 – Delivery of aminoacylated tRNAs translational machinery. (A) Correctly aminoacylated tRNAs will lead to production of correctly folded WT proteins. (B) Misacylated tRNAs will produce errors that induce misfolding, which cause initiation of the unfolded protein response. Adapted from Lee *et al.*, 2006.

High error rate in protein synthesis has been shown to cause disease phenotypes in mouse models where global mistranslation produce tissue-specific neurodegeneration (44, 45). For example, the sticky mutation was identified by Lee and coworkers in mice that exhibited progressive neurodegeneration and cerebellar Purkinje cell loss in the cerebellum. Due to a mutation in the editing domain of alanyl-tRNA synthetase, its proofreading activity was compromised which increased mischarching of non-cognate amino acids into Ala-tRNAs,

consequently leading to global mistranslation of codons decoded by these tRNAs, which led to accumulation of misfolded/unfolded proteins (44). Misfolded proteins are associated with multiple neurodegenerative diseases such as Alzheimer's, Parkinson's and Huntington's diseases, and amino acid misincorporation may be a key trigger in the pathology of multiple sclerosis and amyotrophic lateral sclerosis (46, 47). Nevertheless, in natural conditions, the protein quality control machinery maintains a basal threshold of translational errors at a frequency of around 10^{-4} (46). Molecular chaperones are crucial to these mechanisms as they recognize misfolded proteins, which can either be refolded, degraded via the ubiquitin-proteasome pathway, or delivered to specialized compartments that sequester potentially toxic species (47). In the same way as the sticky mice, the woozy mutation was described with a similar phenotype of Purkinje cell loss (with exception of the cells in the most distal caudal lobule) as result of the accumulation of misfolded/unfolded proteins, due to a disruption in one of the chaperones of BiP, which is essential in misfolded protein translocation and folding, and ER stress sensing (45).

Despite the inherent negative impact on proteome, ambiguous decoding of mRNA can create selective advantages in some circumstances. Reassignment of stop codons to selenocysteine (Sec) and pyrrolysine (Pyl) clearly proves this point. Selenium is an essential dietary micronutrient with antioxidant properties and its biological effects are delivered by selenium-containing proteins, the selenoproteins. These are present in all three domains of life but not in all species. Selenocysteine is the biological form of selenium in proteins, which is inserted at specific UGA codons by a complex selenosome machinery (39, 48). In eukaryotes, a unique Sec-tRNA is serylated by SerRS which is then converted to Sec. Selenocysteine is inserted in response to UGA stop codons in a very complex mechanism coordinated by an extraordinary number of factors (49). Also, in *Euplotes crassus*, specific UGA codons are decoded both as Sec and Cys and such ambiguity is necessary to produce fully functional proteins (50). On the other hand, in the methanogen family *Methanosarcinaceae*, UAG stop codons in specific genes are translated as pyrrolysine. This amino acid is incorporated in methylamine methyltransferases, rendering these methanogenic archaea their unique capability of methane synthesis using methylamines as precursors (51). This demonstrates that codon ambiguity is tolerable and indeed functional.

It has been shown recently in HeLa cells grown under optimal conditions, that methionine misacylation corresponds to ~1.5% of all methionylated tRNAs, unveiling an unexpected basal level of mismethionylation in mammalian cells. Interestingly, upon viral infection the level of tRNA mismethionylation increased ~13%, and further increases were observed upon exposure to ROS-inducing agents, and upon induction of the unfolded protein response. The authors found that the

trigger to this induction was oxidative stress (52), which is in concordance with the protective role conferred by Met residues in proteins against damage mediated by ROS, seen in *E. coli* (53). Noteworthy, mismethionylation has been proved to occur in mammalian (52), yeast (54) and bacterial cells (55). On the other hand, mistranslation induced by severe oxidative stress has been proved to be toxic in *E. coli*. Hydrogen peroxide is capable of oxidizing a critical cysteine residue in the editing domain of ThrRS, leading to serine misacylation of Thr-tRNAs, which induced protein mistranslation and reduced growth rate (56).

In yeast, expression of the mutant *Candida albicans* Ser-tRNA$_{CAG}$ induced highly detrimental ambiguity in the Leu-CUG codon and major decrease in fitness due to high level synthesis of aberrant proteins. These triggered the expression of a stress response that provided important selective advantages under stress conditions and allowed ambiguous cells to survive in otherwise lethal environments (e.g. toxic doses of heavy metals, salts and cycloheximide) (57). Recently, the first microorganism with an altered genetic code was engineered by reversion of the CUG identity in *C. albicans* from serine back to leucine. Surprisingly, these strains adapted to increasing Leu misincorporation, recovered growth rate to wild-type levels and displayed unexpected phenotypic variability, with highly variable colony and cell morphologies, and increased tolerance to antifungals (58). Altogether this indicates that genetic code alterations are not mere abnormalities and can, in fact, represent a potential to adaptation, allowing species to colonize new ecological niches.

1.3 *Saccharomyces cerevisiae* as a biological model

Yeast is a good biological model to study cellular processes conserved in Eukaryotes and its genome can be easily manipulated. It has a rapid growth and generation time, and is particularly easy to manipulate, replicate and maintain. *S. cerevisiae* cells are round and stable in haploid and diploid form. Their size depend on their state, strain and growth phase, and normally ranges from 4 to 10 µm (59).

Yeasts genes are organized in 16 chromosomes and approximately 6,600 open reading frames (ORFs) have been annotated, with more than 80% functionally characterized (59, 60). In 1996, *S. cerevisiae* genome was the first sequenced eukaryotic genome (61). Due to conservation of homologous genes to human, yeasts are used as model systems for the study of human diseases. In fact, 60% of yeast genes have human homologues or at least one conserved domain with human genes (60). In addition, up to 30% of genes implicated in human diseases have a close yeast homologue (62).

There are a vast repertoire of well-established and widely used yeast cellular and molecular techniques, together with modern high-throughput tools (60). Particularly, DNA transformation techniques in yeast proved a powerful useful tool for genomics and proteomics studies as plasmids can be easily introduced in yeast cells either as replicating molecules or by genome integration (59, 63). This proves yeast as a practical and resourceful model organism.

Gene deletion techniques are also easily applicable. One example is the knock-out collection from Bloom-Ackermman and colleagues, used in this thesis, consisting of 204 nuclear encoded tRNA genes deletion, and even some double deletions. Their work exposed the highly complex architecture of the RNA pool, which revealed an extensive network of backup-compensation between and within tRNA families and, interestingly, that genes encoding identical tRNAs within the same family contribute differently to cell fitness. Also, deletions in single-copy and multi-copy tRNA families elicited different transcriptional responses. Of particular interest to this work, they identified two single-copy tRNA families, tL(GAG)G and tR(CCU)J, as non-essential upon deletion. This genetic perturbation resulted in a strong growth defect of the cells due to the inexistence of a cognate tRNA to decode the CUC-leucine codon and AGG-arginine codon (64), rendering them "orphan codons".

1.4 Working hypothesis and objectives

The host laboratory has made a major breakthrough by engineering the first complete codon reassignment in the human pathogen *Candida albicans*, using a combination of tRNA engineering, gene recombination and forced evolution (58). This raised the hypothesis that synthetic codon ambiguities combined with experimental evolution have the power to re-code rare sense codons.

In this project, we used the described above yeast tRNA gene deletion library, particularly the two strains with viable single-copy tRNA gene deletions to test the hypothesis that rare codons can be reassigned using experimental evolution. For this, we engineered a mutant serine tRNA (58) to misincorporate Ser at CUC-leucine and AGG-arginine sites on a proteome wide scale, and evolved the ambiguous strains using experimental evolution. The specific objectives of this Masters Thesis are the following:

1 – Construction of *Saccharomyces cerevisiae* strains with ambiguity in rare Leu and Arg codons;

2 – Experimental evolution of ambiguous strains to reassign sense codons;

3 – Development of a reporter system to monitor Ser misincorporation at CUC-leucine and AGG-arginine sites.

Chapter 2

2. Materials and Methods

2.1 Strains and growth conditions

2.1.1 Strains

Escherichia coli strain JM109 (recA1 SupE44 endA1 hsdR17 gyrA96 relA1 thi Δ[Lac-proAB] F'[traD36 proAB-lacI lacZΔM15) was used as a host for all DNA manipulations.

Saccharomyces cerevisiae strains used in this thesis are haploid and based on the genetic background of Y5565 (table 2.1). ΔtL(GAG)G and ΔtR(CCU)J are knock-out strains that have a deletion of the single-copy genes tL(GAG)G and tR(CCU)J that encode $tRNA^{Leu}_{GAG}$ and $tRNA^{Arg}_{CCU}$, respectively. These strains belong to a tRNA gene deletion library that was kindly provided by Dr. Yitzhak Pilpel (64).

Table 2.1 – Yeast strains used in this study

Strain	Genotype
Y5565	*MATα, can1Δ::MFA1pr-HIS3, mfα1Δ::MFα1pr-LEU2, lyp1Δ, ura3Δ0, leu2Δ0, his3Δ1, met15Δ0*
ΔtL(GAG)G	*ΔtL(GAG)G::Hyg, MATα, can1Δ::MFA1pr-HIS3, mfα1Δ::MFα1pr-LEU2, lyp1Δ, ura3Δ0, leu2Δ0, his3Δ1, met15Δ0*
ΔtR(CCU)J	*ΔtR(CCU)J::Hyg, MATα, can1Δ::MFA1pr-HIS3, mfα1Δ::MFα1pr-LEU2, lyp1Δ, ura3Δ0, leu2Δ0, his3Δ1, met15Δ0*

2.1.2 Growth conditions

E. coli cells were grown at 37 ºC in LB broth medium or LB with 2% agar (Formedium) supplemented with ampicillin (75 µg/ml; Sigma-Aldrich).

Saccharomyces cerevisiae cells were grown at 30ºC in YPD (2% glucose). Transformed *S. cerevisiae* cells were grown in minimal medium lacking uracil (MM-Ura; 0,67% yeast nitrogen base without amino acids, 2% glucose, 2% agar and 100 µg/ml required amino acids – drop-out mixture in annexes).

After evolution experiments, yeast cells were grown in 5-fluoro-orotic acid (5-FOA) plates (description in annexes). The purpose of this assay was to select cells that lost the plasmid containing the URA3 marker, so that transformation of the same cells with another plasmid (with

the same marker) was possible. For this, 100 µl of an evolved clone from each strain were plated in 5-FOA plates and left at 30ºC for 2 days, until colonies were visible.

2.2 Oligonucleotides

Oligonucleotides (table 2.2) were purchased from IDT – Integrated DNA Technologies and ressuspended in miliQ water to a final concentration of 100 mM.

Table 2.2 – List of oligonucleotides used in this thesis

Oligo	Sequence (5' -> 3')	Tm (ºC)
oUA2133	CGCGTCGACGTCCAGGACTGATTTATGTGCATC	60
oUA2134	CGCGGATCCCAGTATGGATTGCTAGTCCTAGAG	60
oUArg1	CGCACTAGTCAGTATGGATTGCTAGTCCTAGAG	61.5
oUArg2	GGTTAAGGAGAAAGACTACGAATCTTTTGGGCTTTGC	62.7
oUArg3	GCAAAGCCCAAAAGATTCGTAGTCTTTCTCCTTAACC	62.7
oUArg4	GGTTAAGGAGAAAGACTTAAAATCTTTTGGGCTTTGC	60.9
oUArg5	GCAAAGCCCAAAAGATTTTAAGTCTTTCTCCTTAACC	60.9
oUArg6	AAGGCGATTAAGTTGGGTA	51.1
oUArg7	ACACAGGAAACAGCTATGA	51
oUArg10	GGTTAAGGAGAAAGACTGAGAATCTTTTGGGCTTTGC	62.8
oUArg11	GCAAAGCCCAAAAGATTCTCAGTCTTTCTCCTTAACC	62.8
oUArg14	CGCACTAGTGTCCAGGACTGATTTATGTGCATC	62.9
oUArg15	CGCGCGGCCGCCAGTATGGATTGCTAGTCCTAGAG	69.9
oUArg16	TATGGTACCCTAGCTTATTTGTACAATTCATC	56.5
oUArg17	TATCTCGAGCTCGAGGAGCTATTAAGATC	58.4
oUArg18	GTTACCAGACAACCATTACCUCTCCACTCAATCTGCCTTAT	64.4
oUArg19	ATAAGGCAGATTGAGTGGAGAGGTAATGGTTGTCTGGTAAC	64.4
oUArg20	GTTACCAGACAACCATTACTCTTCCACTCAATCTGCCTTAT	63.5
oUArg21	ATAAGGCAGATTGAGTGGAGAGGTAATGGTTGTCTGGTAAC	64.4
oUArg22	AGAAGGTTATGTTCAAGAAAGGACTATTTTTTTCAAAGATG	59.2
oUArg23	CATCTTTGAAAAAAATAGTCCTTTCTTGAACATAACCTTCT	59.2
oUArg24	AGAAGGTTATGTTCAAGAATCTACTATTTTTTTCAAAGATG	57.9
oUArg25	CATCTTTGAAAAAAATAGTAGATTCTTGAACATAACCTTCT	57.9
oUArg26	TATGGATGAATTGTACAAA	43.4
oUArg27	CATTCTTTTGTTTGTCAGCC	50.2
oUArg28	GTATTCCAATTTGTGACC	45.6
oUArg29	GGTAAATTGCCAGTTCCATG	51.9
oUArg30	TGTGTGGAATTGTGAGCGGA	56.9

2.3 Plasmids

2.3.1 Original plasmids

In this work we used the following plasmids (table 2.3) as backbone for the new constructions (section 2.3.2):

Table 2.3 – Original plasmids used in this study

Plasmid	Description
pUA261	Plasmid based on pRS315 with a single copy of the *C. albicans* $tRNA^{Ser}_{UGA}$ gene inserted between SalI e BamHI cloning sites
pRS426	Multi-copy plasmid with 5726 bp containing the AmpR and URA3 gene, allowing for selection of transformed cells in LB media with ampicillin or minimal media lacking uracil, respectively
pRS316	Single-copy plasmid with 4887 bp belonging to a series of pBluescript-based centromere vectors. Contains the AmpR and URA3 gene
pACT1-GFP	Plasmid based on CIP10 containing the codon-optimized yeast enhanced GFP (*yE*GFP) gene (65)

2.3.2 Constructed plasmids

For expression of the chimeric *C. albicans* $tRNA^{Ser}_{UGA}$ gene in *S. cerevisiae*, we used the plasmid pUA261 to amplify the $tRNA^{Ser}_{UGA}$ gene with the primers oUA2133 and oUA2134. The gene was then cloned into *SalI* e *BamHI* restriction sites of pRS426 and pRS316, originating the plasmids pUA715 and pUA716, respectively (figure 2.1 in annexs).

The anticodon of the $tRNA^{Ser}_{UGA}$ gene, inserted in pUA715 and pUA716, was altered by site-directed mutagenesis to CCU and GAG in order to generate mutant misreading tRNA genes. Expression of $tRNA^{Ser}_{CCU}$ from multi and single-copy plasmids (pUA717 and pUA718) promotes the insertion of serine at AGG arginine sites in strains lacking the tRNA that recognizes this codon (ΔtR(CCU)J). Expression of $tRNA^{Ser}_{GAG}$ from multi and single-copy plasmids (pUA719 and pUA720) promotes the insertion of serine at CUC leucine sites in strains lacking the tRNA that recognizes this codon (ΔtL(GAG)G).

In order to monitor Ser misincorporation, we developed a loss-of-function reporter system based on the codon-optimized yeast enhanced GFP (yEGFP) gene and assembled it on the multi-copy plasmids pRS426, pUA715, pUA717 and pUA719, in different versions. For that, first we amplified the yEGFP gene plus the promoter ACT1 from pACT1-GFP, using the primers oUArg16 and oUArg17, and then inserted it between the *KpnI* e *XhoI* cloning sites of the referred plasmids. This

originated the plasmids pUA721 (figure 2.2 in annexes), pUA728, pUA733 and pUA736, respectively.

To monitor misincorporation of serine at leucine sites, we used a strategy previously optimized in the host laboratory for *C. albicans* (58) on the plasmids pUA721, pUA728 and pUA733. We changed the leucine codon UUA in position 201 (L201) of the *y*EGFP by site-directed mutagenesis to leucine codon CUC, yielding plasmids pUA725, pUA730 and pUA735. In this way, when Ser is incorporated at residue 201, it causes a destabilization in GFP rendering it inactive, permitting the monitorization of Ser misincorporation by $tRNA^{Ser}_{GAG}$. Next, in order to have a negative control, we altered this same codon to the serine UCU codon, originating plasmids pUA723, pUA729 and pUA734.

Lastly, to monitor misincorporation of serine at AGG arginine sites, we used an analogous strategy on the plasmids pUA721, pUA728 and pUA736. For that, we changed the codon AGA in position 96 (correspondent to arginine, R96) by site-directed mutagenesis to codon AGG, yielding the plasmids pUA727, pUA732 and pUA738. Like in the last case, when Ser is incorporated at this residue renders GFP inactive, allowing the monotorization of Ser misincorporation by $tRNA^{Ser}_{CCU}$. Next we produced the negative control altering the codon of this residue to UCU, originating plasmids pUA726, pUA731 and pUA737.

Table 2.4 – List of the engineered plasmids

Plasmids	Description
pUA715	Multi-copy plasmid with insertion of Ser-$tRNA_{UGA}$ gene between *SalI* e *BamHI*
pUA716	Single-copy plasmid with insertion of Ser-$tRNA_{UGA}$ gene between *SalI* e *BamHI*
pUA717	Plasmid based on pUA715, constructed by site directed mutagenesis of the anticodon of the Ser-$tRNA_{UGA}$ gene to CCU
pUA718	Plasmid based on pUA716, constructed by site directed mutagenesis of the anticodon of the Ser-$tRNA_{UGA}$ gene to CCU
pUA719	Plasmid based on pUA715, constructed by site directed mutagenesis of the anticodon of the Ser-$tRNA_{UGA}$ gene to GAG
pUA720	Plasmid based on pUA716, constructed by site directed mutagenesis of the anticodon of the Ser-$tRNA_{UGA}$ gene to GAG
pUA721	Plasmid pRS426 with insertion of ACT1-GFP between the restriction sites of *KpnI* and *XhoI*
pUA723	Plasmid based on pUA721, constructed by site directed mutagenesis of the codon correspondent to L201 to TCT
pUA725	Plasmid based on pUA721, constructed by site directed mutagenesis of the codon correspondent to L201 to CUC

pUA726	Plasmid based on pUA721, constructed by site directed mutagenesis of the codon correspondent to R96 to TCT
pUA727	Plasmid based on pUA721, constructed by site directed mutagenesis of the codon correspondent to R96 to AGG
pUA728	Plasmid based on pUA715 with insertion of ACT1-GFP between the restriction sites of KpnI and XhoI
pUA729	Plasmid based on pUA729, constructed by site directed mutagenesis of the codon correspondent to L201 to TCT
pUA730	Plasmid based on pUA729, constructed by site directed mutagenesis of the codon correspondent to L201 to CUC
pUA731	Plasmid based on pUA729, constructed by site directed mutagenesis of the codon correspondent to R96 to TCT
pUA732	Plasmid based on pUA729, constructed by site directed mutagenesis of the codon correspondent to R96 to AGG
pUA733	Plasmid based on pUA719 with insertion of ACT1-GFP between the restriction sites of *KpnI* and *XhoI*
pUA734	Plasmid based on pUA733, constructed by site directed mutagenesis of the codon correspondent to L201 to TCT
pUA735	Plasmid based on pUA733, constructed by site directed mutagenesis of the codon correspondent to L201 to CUC
pUA736	Plasmid based on pUA717 with insertion of ACT1-GFP between the restriction sites of *KpnI* and *XhoI*
pUA737	Plasmid based on pUA736, constructed by site directed mutagenesis of the codon correspondent to R96 to TCT
pUA738	Plasmid based on pUA736, constructed by site directed mutagenesis of the codon correspondent to R96 to AGG

2.3.3 Site directed mutagenesis

This procedure was used with the goal of mutating the anticodon of *C. albicans* $tRNA^{Ser}_{UGA}$ gene and the codons of leucine and arginine residues (L201 and R96, respectively) of yEGFP gene. For that, polymerase chain reactions (PCR) were prepared with 1x Pfu Buffer with MgSO4, 0.2 mM dNTPs, 0.2 mM of each respective primer, 1.25 U of Pfu polymerase, 5-25 ng of plasmidic DNA, and miliQ water to a final volume of 25 µl. The PCR program used consisted of a cycle of 30 seconds at 95 ºC, 1 minute at 55 ºC and 7 minutes at 68ºC. Finally, the resulting products were treated with 0.5 µl of *DpnI* for 2 hours, at 37 ºC. The goal was to digest the parental DNA template and select for the mutation-containing DNA. After digestion, plasmids were transformed in *E. coli* competent cells and plasmid DNA was extracted for DNA sequencing (StabVida sequencing services).

2.4 Preparation of *E. coli* competent cells

E. coli competent cells were prepared from strain JM109 using the TFB method (66) . First, 200 µl of an overnight culture was inoculated in 5 ml of LB medium and left in the incubator at 37 ºC, 180 rpm, until they reached an OD_{600} of 0.3. Then, 4 ml of the late culture were inoculated in 100 ml of LB medium and left to grow in the same conditions until they reached an OD_{600} of 0.3. After this, cells were collected in two falcons, cooled on ice for 5 minutes and centrifuged at 2500 rpm for 5 minutes, at 4 ºC. The supernatant was discarded and the pellet ressuspended in 20 ml of cold TFBI (0.03 mM potassium acetate, 0.08 mM $RbCl_2$, 0.013 mM $CaCl_2$, 0.08 mM $MnCl_2$, 15.4% glycerol, pH 5.8). Cell suspension was centrifuged at 2500 rpm for 5 minutes, at 4 ºC and after rejection of the supernatant, the pellet was ressuspended in 2.5 ml of cold TFBII (0.01 mM MOPS Ca, 0.01 mM $CaCl_2$, 0.008 mM $RbCl_2$, 13.4% glycerol, pH 6.5). Cells were then incubated on ice for 5 minutes and distributed on aliquots of 200 µl each and stored at -80 ºC.

2.5 Plasmidic DNA purification from *E. coli*

2.5.1 Miniprep Kit

For the extraction and purification of plasmidic DNA, we used GeneJET™ Plasmid Miniprep kit (Fermentas). Recombinant *E. coli* cells were inoculated in 5 ml or 10 ml of LB medium with 75 µg/ml ampicillin in the case of multi-copy or single-copy plasmids, respectively, and grown overnight at 37 ºC. Cells were harvested at 8000 rpm for 2 minutes, the supernatant was removed and the procedure was followed according to the manufacturer instructions. After the purification, DNA concentration was quantified using the NanoDrop system and plasmids were stored at -20 ºC.

2.5.2 "Homemade minipreps"

For screening procedures, a cheap miniprep protocol that avoided the use of the kit columns was established. Briefly, 1.5 ml of *E. coli* cultures grown overnight were collected by two rounds of centrifugation at 13000 rpm, ressuspended in 1 ml of cold solution I (0.5 mM glucose, 250 mM Tris, 10 mM EDTA) and incubated on ice for 5 minutes. Then, it was added 200 µl of solution II (0.2 M NaOH, 1% SDS), at room temperature, and incubated again on ice for 5 minutes. This was followed by the addition of 150 µl of cold solution III (1 mM KOAc saturated with KOH) and by other incubation on ice. The suspension was then centrifuged at 13000 rpm for 10 min and supernatant

was collected. After a series of precipitations with isopropanol and ethanol 70%, the pellet was dried, ressuspended in water and quantified using NanoDrop.

2.6 Plasmidic DNA purification from *S. cerevisiae*

For the extraction and purification of plasmidic DNA from yeast, we used Wizard® Genomic DNA Purification Kit from Promega, and followed the instructions from the manufacture.

2.7 Transformation procedures

2.7.1 Escherichia coli

Transformation of *E. coli* cells was performed using the chemical SOC method (66). For this, 10 ng of plasmidic DNA was added to thawed aliquots of competent cells (JM109), mixed and incubated on ice for 30 minutes. This was followed by a heat-shock at 42 ºC for 90 seconds and then cooled on ice for 2 minutes. Next, 800 µl of SOC (for preparation of 100 ml at pH 7, 2 g of tryptone, 0.5 g of yeast extract and 0.05 g of NaCl were weighted, and 1 ml of KCl 250 mM and 20 ml of glucose 1M were added) were added to the transformation mix and then incubated at 37 ºC, 180 rpm for 1 hour. After this, cell suspension was centrifuged at 2500 rpm for 1 minute and the supernatant was removed, leaving approximately 50 µl that were homogenized and plated in solid LB supplemented with 75 µg/ml of ampicillin. Plates were incubated at 37 ºC, overnight.

2.7.2 *Saccharomyces cerevisiae*

Yeast transformation was accomplished using the lithium acetate (LiAc) method described by Gietz and Woods (63). Briefly, 1.5 ml of overnight cultures were harvested in log phase by centrifugation for 1 minutes at 4000 rpm. The supernatant was discarded and the following reagents were added to the pellet in sequence: 240 µL of PEG 50% (w/v), 36 µL of LiAc 1 M, 25 µL of previously denatured single-stranded salmon sperm carrier DNA (2 mg/mL) and 36 µl dH_2O plus 0.1-1 µg of plasmid DNA. The mixture was ressuspended and submitted to a heat-shock treatment at 42 ºC for 1 hour. Cells were harvested by centrifugation for 1 minute at 13000 rpm, the supernatant was removed and the pellet resuspended in minimal medium lacking uracil. The cell suspension was then plated in solid MM-Ura medium and incubated for 3 to 5 days at 30 ºC, until colonies were visible.

2.7.3 *Saccharomyces cerevisiae* (alternative)

When the LiAc method proved unfruitful, we used an adapted and modified transformation protocol for *Candida albicans* (67). In this version, 5 ml of overnight cultures were harvested in log phase by centrifugation for 5 minutes at 4000 rpm and the pellet was ressuspended in 150 µl LiAc-sol (10% LiAc 1M, pH 7.5; 10% TE 10x, pH 7.5). 5 µg of plasmid DNA was added to the suspension, followed by 100 µg of single-stranded DNA carrier and 600 µl of PEG/LiAc-sol. (50% PEG 50% (w/v), 50% LiAc-sol), and mixed. The transformation mix was incubated overnight at 30ºC, 180 rpm and submitted to a heat-shock at 44 ºC during 15 minutes and then cooled on ice about 2 minutes. The cell suspension was centrifuged in a bench top centrifuge at maximum speed for 1 minute and the supernatant was discarded. The pellet was ressuspended in 150 µl of MM-Ura and platted in solid MM-Ura medium and incubated for 3 to 5 days at 30 ºC, until colonies were visible.

2.8 Colony PCR

Colony PCR was used to check for the correct transformation of *E. coli* and yeast cells. For this, individual colonies from selective plates were picked using a P10 tip and were homogenized in 5 µl of miliQ water. Cell suspension was submitted to a heat shock at 95 ºC for 5 min. Next, PCR reagents were added to each tube: 1x DreamTaq Buffer, 0.2 mM dNTP mix, 0.15 mM of each primer, 1.25 U of DreamTaq polymerase and miliQ water to a final volume of 15 µl. General PCR programs consisted an initial denaturation at 95 ºC for 3 minutes, followed by 30 cycles of a denaturation step at 95 ºC for 30 seconds, a annealing step for 30 seconds at variable temperature and an extension step at 72 ºC for 1 minute. Reactions were performed in a MyCyclerTM thermal cycler (BIORAD) and ended with a final extension step at 72 ºC for 5 minutes. The PCR products were subjected to a gel electrophoresis for quality check.

2.9 Growth curves and rates

Growth curves were established for all strains in minimal medium lacking uracil or YPD in the case of non-transformed cells. First, cultures were pre-incubated for 3 days in the appropriate medium at 30ºC with constant shaking at 180 rpm. The optical density (OD) of the pre-inoculum was measured at 600 nm and then the cultures were inoculated in fresh media to an initial OD_{600} of 0.01. After approximately 7 hours, the OD was monitored each hour until they reached stationary phase. Data obtained was plotted in a graph as $\log(OD_{600})$ as a function of time (h). A trend line was

a fitted to the exponential phase of growth and its slope represented the growth rate of the clones. This procedure was carried out with five different clones from each strain, in the non-evolved and evolved state.

2.10 Evolution experiments

Experimental evolution assays were carried out using 5 clones of each newly constructed strain, as well as wild-type and knock-out control strains. Control cells were grown in 1 ml of YPD, whereas transformed cells were grown in 1 ml of minimal medium lacking uracil at 30ºC, until they reached stationary phase. Then cells were diluted by a factor of 1:200 into fresh medium (number of generations). This procedure was repeated until the strains reached approximately 200 generations. Number of generations was calculated as shown in the equation below.

$$N_0 = N_t / 2^g$$

N_0 = Initial optical density
N_t = Final optical density
g = Number of generations

2.11 Northern Blot Analysis

2.11.1 Total RNA extraction

For total RNA extractions, 50 ml cultures of *S. cerevisiae* cells were grown overnight in minimal medium and harvested at an OD600 of 1-1.5. The pellets were then frozen overnight at -80ºC.

Frozen cells were resuspended in hot acid-phenol:chloroform 5:1 (pH 4.7) and TES-buffer (10 mM Tris pH 7.5, 10 mM EDTA, 0.5% SDS) and incubated for 1 hour, at 65ºC, with repeated shaking every 10 minutes. Cell suspension was centrifuged at 13000 rpm for 20 minutes, at 4ºC. Then the RNA containing aqueous phase was transferred to new tubes and the same volume of acid-phenol:chloroform 5:1 (pH 4.7) was added. The suspension was mixed using a vortex and centrifuged at 13000 rpm for 10 minutes, at 4ºC and this step was repeated. The aqueous phase was transferred to a new tube with the same volume of chloroform:isoamil-alchool 24:1, vortexed hard and centrifuged at 13000 rpm for 10 minutes, at 4ºC. Then 350 µl of the aqueous phase was transferred to a new tube containing 35 µl of sodium acetate (3 mM NaAc, pH 5.2) and 800 µl of ethanol 100% (-20ºC). The aqueous phase was precipitated overnight at -20ºC. Tubes were then centrifuged at 13000 rpm for 5 minutes, at 4ºC. The fluid was carefully removed without touching

the RNA-pellet and the later was washed with 500 µl of ethanol 80% (-20ºC) and briefly spinned down at 13000 rpm. The ethanol was removed and the tubes were air-dried so that all traces of ethanol were removed. The pellet was dissolved in 100 µl of sterile miliQ water and the concentration was quantified by Nanodrop.

2.11.2 Northern blot

Fractionation of total tRNAs was carried out on 12-15% polyacrylamide (40% Acril:Bis) gels containing 8 M urea (0.8 mm thick, 30 cm long). In each gel slot, 50 µg of total RNA sample was loaded and gels were electrophoresed at 500 V for 16 hours. Fractioned tRNAs were localized by UV shadowing, the portion of the gel containing tRNAs was cut and transferred onto a nitrocellulose membrane (Hybond N, Amersham) using a Semy-dry Trans Blot (Bio-Rad). For hybridization, probes were prepared using 10 pmol of dephosphorylated oligonucleotide and 4 µl of γ-^{32}P-ATP (5000Ci/mmol) (Perkin Elmer) in 1x T4 kinase buffer, 10 mM spermidine and 16 units of T4 kinase (Takara). Labelling reaction was incubated at 37ºC for 1 hour and then the probe was extracted using phenol:chloroform:isoamyl alcohol (PCIA). The hybridization protocol was performed as described by Jacques Heitzler (68). Membranes were pre-hybridized for 30 minutes at 53 ºC in a hybridization solution [5x Denhardt's solution (1% Ficol, 1% Polyvinylpyrrolidone and 1% Bovine serum albumin), 6x SSPE (3 M NaCl, 173 mM NaH2PO4, 25 mM EDTA) and 0.05% SDS]. Membrane hybridization was performed overnight in the above buffer using probes oUA2199 for detection of WT and mutant Ser-tRNA and oUA2195 for detection of the WT control Gly-tRNA$_{CCC}$ (table 2.5). Membranes were then washed 4 times (3 minutes each time) in 2x SSPE, 0.5% SDS at 53 ºC. The membranes were exposed overnight with intensifying screens and developed using a Molecular Imager FX (Biorad).

Table 2.5 – List of oligonucleotides used for northern blot analysis

Probe	Detection	Tm (ºC)	Sequence (5' -> 3')
oUA2199	WT and mutant Ser-tRNA$_{UGA}$	51,5	TTAACCGCTCGGACAAGTT
oUA2195	WT Gly-tRNA$_{CCC}$	52-55	GCGGAAGCCGGGAATCGAAC

2.12 Epifluorescence microscopy

To monitor Ser misincorporation, *y*EGFP expression was visualized in *S. cerevisiae* cells by epifluorescence microscopy. Strains were grown overnight in liquid media at mid-exponential phase and aliquots were spotted onto microscope slides. Fluorescence was detected using a Zeiss MC80 Axioplan 2 light microscope, equipped for epifuorescence microscopy with filter set HE38. Photographs were taken using an AxioCam HRc camera and images were analysed using ImageJ software. Mean fluorescence intensities (± standard deviation) were quantified in individual Leucine mutant cells expressing the reporter yEGFP Leu-UUA_{201} (positive control), Ser-UCU_{201} (negative control) and Ser/Leu-GAG_{201} (reporter). In Arginine mutants, intensities were quantified in individual cells expressing the reporter yEGFP Arg-AGA_{96} (positive control), Ser-UCU_{96} (negative control) and Ser/Leu-AGG_{96} (reporter). yEGFP fluorescence (intensity/pixel) was determined for at least 1000 cells in each case.

Chapter 3

3. Characterization of codon ambiguities

3.1 Overview

Despite the highly detrimental proteome chaos that genetic code alterations create, they are known to occur across the three domains of life which invalidate the hypothesis of a non-mutable code. This raises the intriguing question of how can cells cope with changes in the meaning of the code, since 10 to 50% of amino acid substitutions have the potential to disrupt protein function (69), by synthesis of aberrant, misfolded or non-functional proteins. However, genetic code ambiguity is also a source of protein innovation and phenotypic diversity (29, 30, 48, 58, 70).

Generation of new proteins with novel functions for biotechnological, pharmaceutical, green chemistry and biotechnological industries is one of the main goals of Synthetic Biology. Genetic code engineering has gained a lot of interest with the promise of expanding the code with the incorporation of unnatural amino acids into proteins with novel properties. Engineering of novel proteins with artificial non-canonical amino acids has been accomplished by two *in vivo* strategies: supplementation based incorporation method (SPI) and stop codon suppression approaches (SCS) (figure 3.1) (71).

Several studies from Schultz and colaborators, have demonstrated that bacterial, fungal and mammalian cells are highly tolerant to incorporation of artificial amino acids at stop codons, using orthogonal tRNA-synthetase pairs that do not participate in conventional translation (72-74). In *Escherichia coli*, insertion of an orthogonal tyrosyl-tRNA synthetase (TyrRS) from *Methanococcus jannaschii* and a mutant tyrosine amber suppressor tRNA, allows the incorporation of synthetic amino acid O-methyl-L-tyrosine into proteins, in response to the amber codon UAG. Since the orthogonality prevents the new tRNA-synthetase from aminoacylating endogenous tRNAs and the new tRNA from being recognized by the host tRNA-synthetases, the system is highly specific (75). Anderson and co-workers were also able to engineer an orthogonal aaRS-tRNA pair derived from archaeal Lys-tRNA, which efficiently and selectively incorporated a non-canonical amino acid into proteins in response to the quadruplet codon AGGA (76). Genetic code engineering has also been achieved using an editing defective aaRS. Deletion of the editing domain of an isoleucyl-tRNA synthetase (IleRS) resulted in ambiguous translation of isoleucyl codons with non-canonical amino acids (77). These examples illustrate the efficiency of stop codon suppression (SCS) strategies (figure 3.1c).

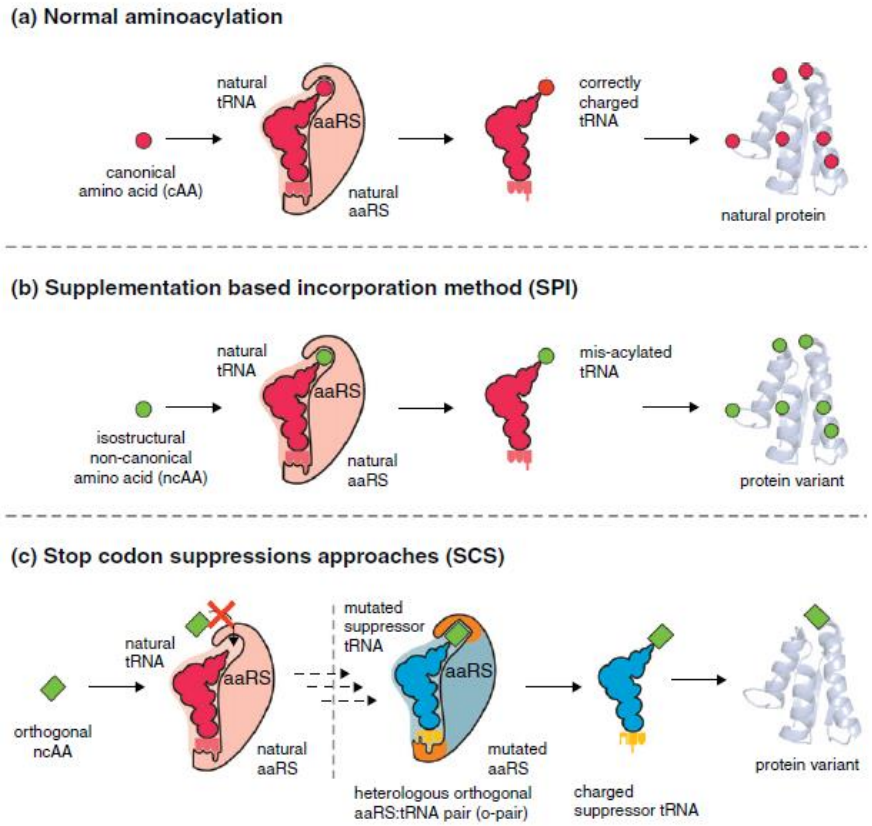


Figure 3.1 Aminoacylation with canonical (a) and non-canonical amino acids for protein translation by supplementation base incorporation (b) and stop codon suppression (c) techniques. Adapted from Hoesl and Budisa, 2012

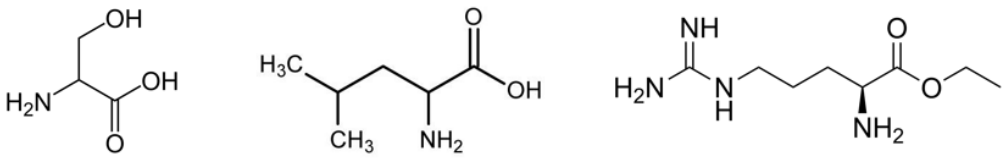
On the other hand, supplementation based incorporation (SPI) strategies use auxotrophic strains and supplements the media with an isostructural non-canonical amino acid so that is not edited by aaRS. In this way, it is possible to exchange several residues in a protein by sense-to-sense reassignments, since it means the incorporation of artificial analogs (71). Indeed this methodology has been used, for example, to engineer mutant *E. coli* that incorporate selenomethionine instead of methionine (78), to engineer fluoropolymers using the non-canonical amino acid p-fluorophenylalanine (79), and the incorporation of labeled methionine analogs into proteins to allow structural and functional protein studies (80). However, due to the low efficiency and the inability to be site-specific, the results obtained from these approaches were rather disappointing and engineering of codon reassignments remains a significant experimental challenge (71).

In this work, we used an alternative strategy of reassignment, similar to the one used by Geslain and coworkers for incorporation of serine at 19 different codons in chick embryos and human cells (81). We engineered ambiguous strains that misincorporate Ser at CUC-Leu and AGG-Arg codons in

a proteome wide fashion, using chimeric tRNAs. As SerRS does not interact with the anticodon of seryl-tRNAs, these can carry mutations in the anticodon as they do not affect their aminoacylation (81). Mistranslation induced on our model system involves the substitution of amino acids with different chemical properties, as highlighted in table 3.2 and figure 3.1, and it is expected to have a negative impact on cell fitness. In fact, according to the amino acid substitution matrix BLOSUM 62 (82), replacement of leucine and arginine for serine has a substitution value of -2 and -1, respectively, which means these are, in fact, unfavoured events.

Table 3.1 Principal characteristics of amino acids involved in this study

	Serine	**Leucine**	**Arginine**
Polarity	Polar	Nonpolar	Positive
Side chain functional group	-OH	Methylene	Methylene
Molecular weight (D)	87	113	156
Isoelectric point	5.7	6	10.8
Polar requirement **(83)**	7.5	4.9	9.1
Side chain flexibility	Low	Moderate	High



Serine **Leucine** **Arginine**

Figure 3.2 Structure of serine, leucine and arginine

3.2 Results

3.2.1 Expression of mutant Ser-tRNAs in *S. cerevisiae*

Since SerRS does not interact with the anticodon of tRNAs upon aminoacylation (84), this allows alterations in the anticodon of Ser-tRNAs without affecting their serylation. This makes *C. albicans* Ser-$tRNA_{UGA}$ a perfect tool for this project, as we can mutate its anticodon and thus reprogram it to misincorporate serine at the CUC-Leu and AGG-Arg codons lacking their cognate tRNAs. We chose this chimeric tRNA for the present study, as it does not contain introns and has 14 mismatches with the homologous $tRNA_{UGA}$ from *S. cerevisiae,* allowing for posterior detection of the mutant tRNAs by northern blot analysis, when expressed in *S. cerevisiae*.

So, mutant tRNAs were engineered by anticodon mutagenesis, generating Ser-tRNA$_{GAG}$ and Ser-tRNA$_{CCU}$, and expressed in knock-out strains ΔtL(GAG) and ΔtR(CCU), respectively (figure 3.3), thus generating ambiguous strains. Mutant tRNAs were expressed in single and multi-copy plasmids (denoted as SC and MC, respectively), conferring two degrees of pressure when in competition for the targeted codon. In the Leucine set (strains based on ΔtL(GAG), figure 3.3, left panel), the unique Leu-tRNA$_{GAG}$ gene was deleted and so the three gene family tL(UAG) represent the only tRNA capable of decoding CUC codons with a relative strong codon-anticodon association (64). *S. cerevisiae* Leu-tRNA$_{UAG}$ is able to decode the four Leu-CUN codons by a superwobble mechanism (85), which is associated with its rare unmodified U$_{34}$ (86). Analogously, the plastid Gly-tRNA$_{UCC}$ has an unmodified U$_{34}$ and is also capable of decoding the four GGN-Gly codons, however it suggests that superwobbling decreases mRNA translational efficiency due to delayed recognition of U/C-ending codons (87). So, in this work system, one can expect that mutant Ser-tRNA$_{GAG}$ will have the upper-hand in codon recognition competition, since CUC decoding by Leu-tRNA$_{UAG}$ is theoretically delayed and thus less efficient.

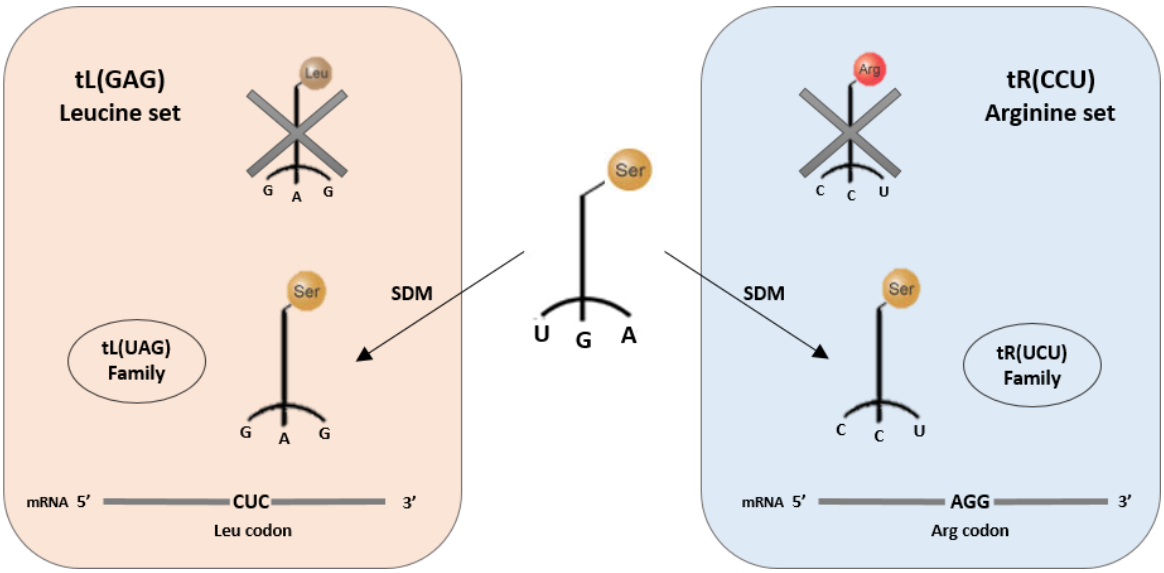


Figure 3.3 – Representation of mutant tRNAs in the strain background of this work. A series of misreading tRNAs were engineered from Ser-tRNA$_{UGA}$ by site directed mutagenesis of its anticodon to GAG and CUC, misincorporating serine at Leu and Arg sites respectively. These codons do not have their cognate tRNAs and their decoding is being compensated by tL(UAG) and tR(UCU) families, respectively

On the other hand, in the Arginine set (strains based on ΔtR(CCU), figure 3.3, right panel), AGG-Arg codon is decoded by Arg-tRNA$_{UCU}$ in the absence of Arg-tRNA$_{CCU}$, by standard wobble rules (64). Its anticodon position 34 is methylated (mcm^{5}U) by Trm9, which enables decoding of G-ending codons by Arg-tRNA$_{UCU}$ (88). It has been shown that AGG decoding by Arg-tRNA$_{UCU}$ is especially

inefficient and causes translational pausing that promotes frameshift +1 (89, 90). Again, in codon competition, it is expectable the advantage of mutant Ser-tRNA$_{CCU}$ since codon-anticodon interaction is stronger due to the G-C pair formed in the wobble position, instead of the less stable mcm^{5}U-G pair formed by Arg-tRNA$_{UCU}$ (89).

Along with the mutant tRNAs, yeast strains were also transformed with the empty multi-copy vector pRS426 (denoted as EV) and the original WT Ser-tRNA$_{UGA}$. Subsequently, five clones from each transformation were selected. To assess the impact of the mistranslating tRNAs in *S. cerevisiae* fitness, growth curves parameters were evaluated.

3.2.2 Leucine set

As expected, different expression levels of the misreading tRNA had different fitness costs to the cell, as seen in figure 3.4. Also, the deletion strain ΔtL(GAG) showed a growth profile similar to the wild-type Y5565, as previously reported (64, 91). There was a strong negative impact in cells highly-expressing Ser-tRNA$_{GAG}$ (ΔtL + MC-tRNA$_{GAG}$) whereas its single-copy counterpart (ΔtL + SC-tRNA$_{GAG}$) had little impact in growth rate (figure 3.4B). Surprisingly, overexpression of the endogenous Ser-tRNA$_{UGA}$ at high level exerted a similar decrease. Indeed, in the two highly expressing strains (ΔtL + MC-tRNA$_{UGA}$ and + MC-tRNA$_{GAG}$), growth rates dropped to around 55% relative to the wild-type strain (figure 3.4B).

Misincorporation of serine at Leu sites involves substitution of chemical distinct amino acids, as highlighted in table 3.1, and so it was expected the subsequent cost in cell fitness. Oddly, this type of mistranslation in a low expressing system (ΔtL + SC-tRNA$_{GAG}$) did not had impact in growth when compared to the deletion strain (ΔtL(GAG)). Bloom-Ackermman and colleagues reported that strains with deletions of single-copy tRNA genes, namely ΔtL(GAG) and ΔtR(CCU), have reduced sensibility to proteotoxic agents. This led them to hypothesize that these cells are already under chronic misfolding stress, which in turn offers cross-protection to further extrinsic proteotoxic stress (64). This may be the reason for the lack of observable phenotype in growth rate of the single-copy group, as the cells may be “programmed” to better respond to further proteotoxic stress, as the incorporation of an erroneous amino acid in a proteome-wide manner. Our results show that this may be true to some extent, as the over expression of the mutant tRNA abolishes this “protection” phenotype.

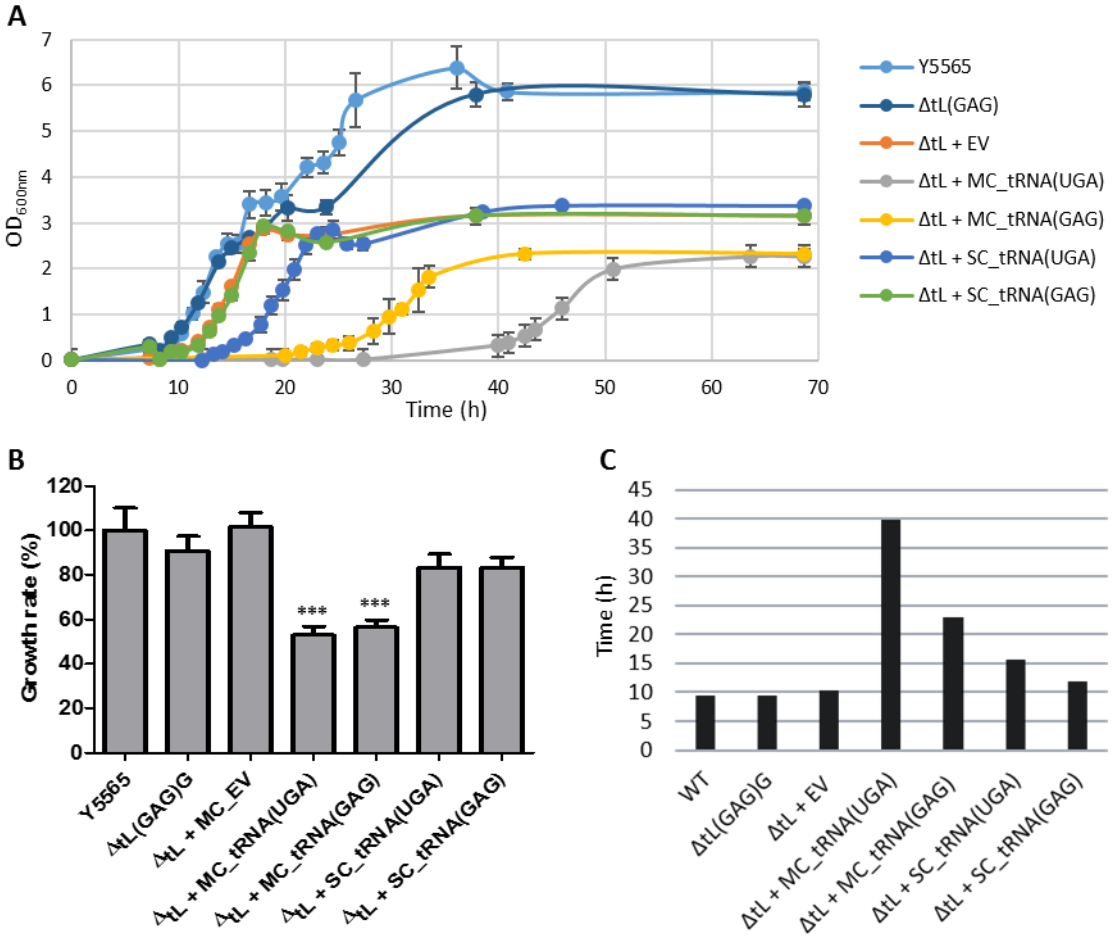


Figure 3.4 – Impact of mutant tRNAs on cell fitness of ΔtL(GAG) mutants. (A) Growth curve measurements represented as optical density (OD) units over time. (B) Growth rate of leucine mutants. Data represents the mean + SD of 5 different clones. ***P <0.001, one-way ANOVA post-Dunnet comparison test with 95% confidence interval relative to ΔtL(GAG) cells. (C) Lag phase duration of leucine mutants, in hours.

One striking feature that arose from growth curves profiles (figure 3.4A) were the drastic differences in the duration of lag phase (figure 3.4C). The duration of a lag phase is indicative of the time needed to reprogram transcription, and so adapt to new challenging conditions (92). Also, the extended lag phase duration indicates that these strains may have defects in metabolism or in cell cycle progression (93).

3.2.3 Arginine set

As previously reported (64, 90), deletion of the gene encoding Arg-tRNA$_{CCU}$ has a significant negative impact in growth rate (figure 3.5) when compared to the wild-type Y5565. However, the Arginine set does not show the same behavior as the Leucine mutants. Also, results from this set

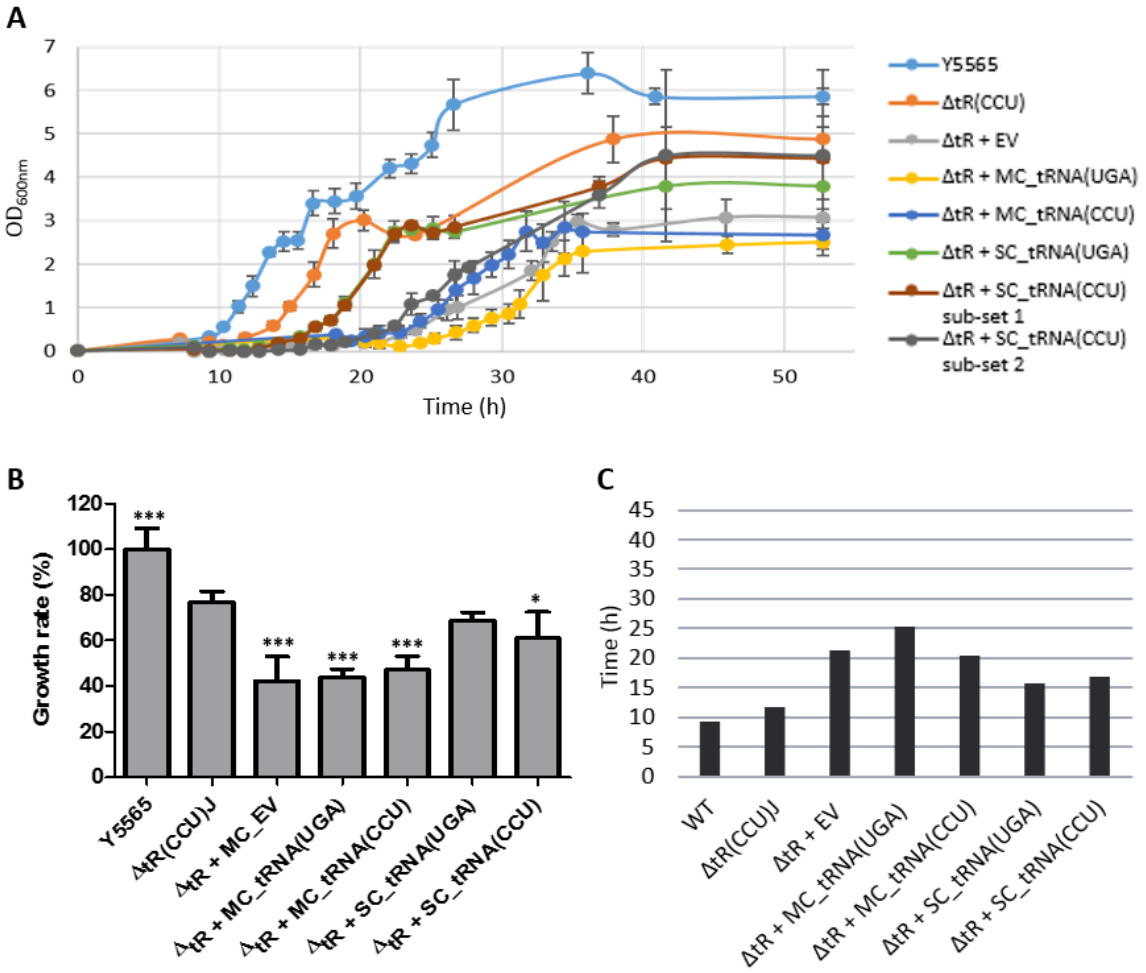


Figure 3.5 – Impact of mutant tRNAs on cell fitness of ΔtR(CCU) mutants. (A) Growth curve measurements represented as optical density (OD) units over time. (B) Growth rate of arginine mutants. Data represents the mean + SD of 5 different clones. ***P <0.001; *P <0.05 , one-way ANOVA post-Dunnet comparison test with 95% confidence interval relative to ΔtR(CCU) cells. (C) Lag phase duration of arginine mutants, in hours.

should be taken carefully and with some reserves, since the strain transformed only with the empty plasmid (ΔtR + EV) is one of the major growth incapacitated strains (figure 3.5A and B). This odd feature was confirmed with three more new independent transformations. We also tested the effect of transforming the empty vector pRS426 in other strains to determine if this phenotype was intrinsic to the knock-out strain ΔtR(CCU). Transformation with the empty vector was performed in two other deletion strains of Leu and Arg-tRNA genes, ΔtL(UAG)J and ΔtR(UCU)M2, respectively. The tL(UAG)J gene encodes one of the copies of the three gene family $tRNA^{Leu}_{UAG}$, while tR(UCU)M2 encodes one of the copies of the eleven gene family $tRNA^{Arg}_{UCU}$, and was characterized as a "major copy", by Bloom-Ackermman, since its deletion had a significant growth impact in comparison with other copies from the same family (64). The growth rate of all these was not altered by the insertion of the plasmid, so the reported result must be due to some constraints in the ΔtR(CCU) strain. One hypothesis to explain this result might be related to the selective marker of the vector, URA3. This

gene encodes the OMP (orotidine-5'-phosphate) decarboxylase, which is involved in uracil synthesis, and its Arg-247 residue is encoded by the codon AGG (94). So, one possible explanation is that the decoding of this codon is highly affected by the absence of its cognate tRNA, in a way that would render part of the enzyme population as not functional. If so, cells would show huge growth defects in selective media lacking uracil, like the case, as they are defective in its production, and consequently represent a situation of nutrient starvation. Consequently, the growth profile shown in figure 3.5 may represent a situation of nutrient starvation. Unfortunately, there is no evidence to corroborate this hypothesis as further studies would be needed.

Hereupon, similarly to the Leucine set, strains expressing the mutant Ser-tRNA$_{CCU}$ (ΔtR + MC-tRNA$_{ccu}$) and Ser-tRNA$_{UGA}$ (ΔtR + MC-tRNA$_{UGA}$) at high level exhibited the highest impact on growth rate, corresponding to a decrease in ~65% relative to WT cells (figure 3.5B). Although there was a visible impact in the duration of the lag phase (~10 more hours than the WT and knock-out strains) (figure 3.5C), the impact of MC-tRNA$_{UGA}$ was less pronounced that in its counterpart in the Leucine set (figure 3.4C).

Another difference is the heterogenous impact of Ser-tRNA$_{CCU}$ in strains expressing it at low level. As seen in figure 3.5A, the behavior of the five clones was different, which led to their division in two sub-sets (ΔtR + SC-tRNA$_{UGA}$). Three clones belong to the first sub-set that was less affected (~30% decrease in growth rate relative to WT) and the two other clones were affected to an extent similar to when expressed in high level (~50% decrease in growth rate relative to WT). Altogether, these results indicate that ΔtR(CCU) strain has a different way to cope with mistranslation that ΔtL(GAG) strain.

3.2.4 Impact of CUC and AGG ambiguity

The results presented above shows that ambiguous strains tolerate the misincorporation of Ser at Leu-CUC and Arg-AGG codons, although with high negative impact to cell fitness. Therefore, to better understand the impact of ambiguous decoding of these codons, and to obtain a picture of its effect on *S. cerevisiae* biology, we studied the genomic distribution and the usage of these codons using the ANACONDA software developed by the Bioinformatics group of Aveiro (95). All the analysis were carried out using the reference genome sequence of *Saccharomyces cerevisiae* strain S288C, available at the *Saccharomyces* Genome Database (SGD, http://yeastgenome.org/).

Analysis with ANACONDA identified 6689 genes in the *S. cerevisiae* genome, and codon distribution analysis revealed that its genome contains 17499 CUC codons distributed throughout

77% of its genes (figure 3.6A). Leu-CUC usage frequency is only 5.5‰ and, therefore, belongs to the category of rare codons. Indeed, most of yeast genes (64.8%) has between 1 to 5 codons, and only 2.4% has more than 10 CUC codons on their ORFs. One the other hand, Arg-AGG usage frequency is a bit higher but still only scores 9.2‰, therefore is still considered a rare codon in yeast (96). Here, ANACONDA codon distribution analysis revealed that its genome contains 28483 AGG codons distributed throughout 84% of its genes (figure 3.6B). Like as CUC, the majority of its genes (56.2%) has 1 to 5 codons. But contrary to CUC distribution, 8.4% of yeast genes has more than 10 AGG codons.

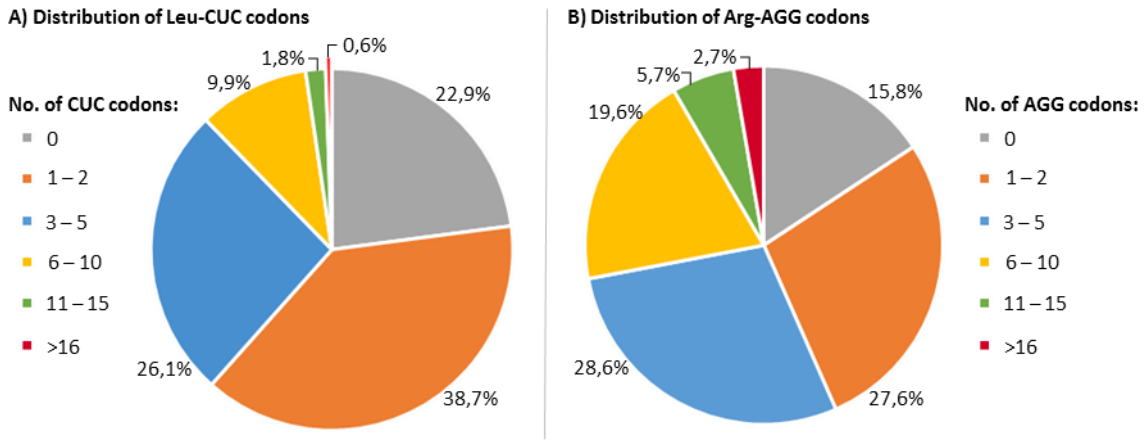


Figure 3.6 – Leu-CUC (A) and Arg-AGG (B) codon distribution over *S. cerevisiae* genome. In the yeast genome almost one fourth of its genes do not have CUC codons, while almost one sixth do not have AGG codons. The majority of its genes, 64.8% in the case of CUC and 56.2% in the case of AGG, contains 1 to 5 codons (orange and blue groups), while 9.9% and 19.6% of its genes have 6 to 10 CUC and AGG codons, respectively (yellow group), and only 2.4% of its genes have 10 or more CUC codons, while in the case of AGG, 8.4% of its genes have 10 or more CUC codons (green and red groups).

Apart from the major decrease seen in growth rate of strains with high expression of mutant Ser-tRNAs, one of the most notorious and unexpected phenotypes was the major increase in the duration of the lag phase. Ma and colleagues reported that *S. cerevisiae* with defective cAMP-dependent protein kinase activity had slow glucose uptake and extended lag phases (93). The cAMP-protein kinase A (PKA) pathway in yeast plays a major role in the control of metabolism, stress resistance and proliferation and so, we looked for *S. cerevisiae* genes related to PKA in the National Center for Biotechnology Information (NCBI) genes database which retrieved a group of 52 genes (complete table 3.1 in annex). Analysis of this group with ANACONDA revealed a different tendency in CUC codon distribution from that in the genome. The majority of these genes have 1 to 5 CUC codons (59.6%) like previously, but the frequency of genes with 6 or more CUC codons increased 10%. Interestingly, almost 10% of this group of genes has more than 10 CUC codons

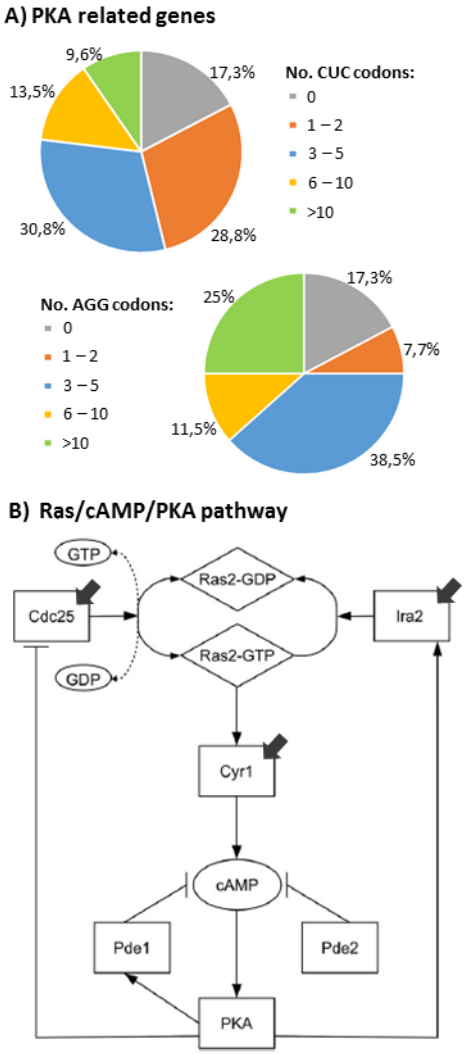


Table 3.2 – PKA related genes with 6 or more CUC codons (and AGG codon content)

Gene	Description	No. codons	No. CUC codons	No. AGG codons
IRA2	GTPase-activating protein	3080	22	31
IRA1	GTPase-activating protein	3093	19	30
CYR1	Adenylate cyclase	2027	16	25
RIM15	Protein kinase involved in cell proliferation	1771	15	21
TOR1	PIK-related protein kinase	2471	13	23
CCC2	Cu^{2+}-transporting P-type ATPase	1005	10	6
GEX2	Proton:glutathione antiporter	616	10	4
GPR1	Plasma membrane G-protein coupled receptor	962	10	5
TOR2	PIK-related protein kinase	2475	10	37
GEX1	Proton:glutathione antiporter	616	9	4
CDC25	Ras-guanine exchange factor	1590	6	20
ROM2	GDP/GTP exchange factor for Rho proteins	1357	6	16

Figure 3.7 – CUC and AGG codon distribution over the PKA related genes group (A) and a schematic of the Ras/cAMP/PKA pathway (B) (adapted from Pescini *et al.*, 2012). (A) Contrary to the tendency in the genome, 23% of this group has more than 6 CUC codons in their genes, while 36.5% has more than 6 AGG (yellow and green groups). (B) Proteins highlighted in the Ras/cAMP/PKA pathway were identified with high number of CUC and AGG codons, as shown in table 3.2.

(figure 3.7A), which is discrepant from the frequency seen in the genome (figure 3.6). Genes from this group show a distribution of AGG codon content even more discrepant. Here, genes with 1 to 5 AGG codons (46.2%) have almost the same frequency as genes with more than 6 codons (36.5%), while one fourth of them has more than 10 codons.

So it is likely that the uncovered group of genes with high number of CUC and AGG codons (particularly those with 10 or more codons) are specially affected by Ser misincorporation at these sites. Table 3.2 shows the identified genes with more than 6 CUC codons, their respective number of AGG codons, the number of codons in their ORFs and their respective biological description. As results obtained from the Arginine set are rather dubious due to the expression of the empty vector,

we focused particularly in CUC-Leu codons, although the same thought can be extrapolated to the Arginine set due to the high number of AGG codons accounted in these genes. Crucial genes involved in the Ras/cAMP/PKA pathway were identified (figure 3.7B and table 3.2), like CYR1 (with 16 CUC codons) which encodes the adenylate cyclase that synthetizes cAMP and therefore activates PKA (97), IRA2 and its paralog IRA1 (with 22 and 19 CUC codons, respectively) which encodes GTPase activating proteins that negatively regulate the activity of Ras proteins (98), and CDC25 (with 6 CUC codons) which encodes a guanine nucleotide exchange factor that positively controls Ras activity by stimulation of GDP–GTP exchange (99). Therefore, this metabolic pathway is probably impaired in our mutant strains and they are expected to have lower levels of intracellular cAMP than WT and deletion strains. So, although PKA itself is not affected by CUC codon content (catalytic subunits TPK1/2/3 have none or one CUC, while its regulatory subunit BCY1 has 5 CUCs, as seen in in table 3.3 in annex), the impaired supply of cAMP may compromise PKA activity which is dependent of cAMP. Low PKA activity has been shown to stop growth in yeast by arresting cells in a G_1 quiescent state, similar to G_0 in higher eukaryotes (100), which is in accordance with our results from extended lag phases, but differs from our mutant strains as they were not permanently arrested.

3.3 Discussion

Mistranslation is expected to be toxic to the cell and to have a deleterious effect in cell growth rate. For example, mistranslation at CUG-Leu codons, in *S. cerevisiae* induced by the expression of the *C. albicans* Ser-$tRNA_{CAG}$ caused a decrease of around 50% in growth rate and triggers the stress response. It is noteworthy that this pronounced impact is a result of the tRNA expression in a single-copy plasmid. (57). So, in our low expression system, one should expect a strong decrease in growth rate due to mistranslation, which was not the case. In our Leucine system, low expression of misreading tRNAs did not have an impact in fitness, when compared to the knock-out strain, which is also true for part of the ΔtR clones (subset 1 of single-copy $tRNA_{CCU}$). These results can be explained by the fact that the knock-out strains used in this study are already in a state of chronic stress, which rendered them less sensitive to proteotoxic agents (64). Microarray data showed that pathways responsive to proteotoxic stress, such as the proteasome and protein processing in ER, are up-regulated in these strains (64). So, since cells have already the stress pathways activated, they are able to better cope with a further proteome destabilizing agent, such as the low expression of our mutant tRNAs.

On the other hand, the arginine single-copy mutants showed a pronounced heterogeneous growth rate and lag phase duration. Organisms respond to environmental changes by adapting the expression of key genes. However, such transcriptional reprogramming is energetically costly and decreases cell fitness (92). The heterogeneity observed can indicate that cell may be "experimenting" different metabolic reprogramming, like seen in response to changing carbon environments (92).

Howsoever, the overall behavior of arginine mutants was different from the one exhibited by leucine mutants. Apart from the stronger decrease in growth upon gene deletion, the microarray data collected by Bloom-Ackermman and colleagues, shows that mRNA expression of ΔtR(CCU) is much more deregulated than in ΔtL(GAG) (64), possibly rendering the first more sensible to further aggression. But contrary to this notion, ΔtR(CCU) shows high tolerance to proteotoxic agents like DTT (reducing agent), AZC (toxic analog of proline) and tunicamycin (induces the unfolded protein response in the ER) (64). However, in the un-engineered ΔtR(CCU) strain, AGG decoding is being compensated by an iso-acceptor tRNA which inserts arginine at low efficiency, so the overall effects are not comparable to our system, since these codons are being decode as serine. So, one hypothesis is that AGG codons are present in specific sites that are less tolerant to point amino acid substitutions.

In both groups, strains expressing the mutant tRNA at high level had a strong negative impact on growth rate and on the time required to enter the exponential phase, but remained viable. Extended lag phases have been reported in *Saccharomyces cerevisiae* with deficient cAMP-dependent protein kinase activity (93). Accordingly, analysis of CUC and AGG codon content in PKA related genes revealed a trend that culminates in lowering PKA activity, particularly by possible disruption of the adenylate cyclase responsible for cAMP synthesis (CYR1). In *S. cerevisiae*, a normal activity of PKA is required for growth and cell cycle progression, while a reduction of its activity, related to a decrease of intracellular cAMP, results in the accumulation of storage carbohydrates, high stress tolerance and arrest of growth and cell cycle (100). Several genes with high frequency of these codons are directly or indirectly related to the cAMP pathway and have high biological relevance. For example, GPR1 is a membrane receptor, part of the glucose-sensing system together with GPA2, which regulates cellular cAMP level by stimulation of cAMP synthesis (101), which also suggests that glucose sensing may be impaired in our mutants.

Interestingly, major destabilizations in the endogenous tRNA pool accomplished by highly expressing the endogenous Ser-tRNA$_{UGA}$, also had a strong impact in cell fitness. In fact, these strains were the most affected and exhibited the longest duration in lag phase. For example, the

Leu mutant (ΔtL + MC-tRNA$_{UGA}$) took an astonishing 40 hours to adapt to their new genetic condition and so enter in the exponential phase (figure 3.4C). One can rationalize that as cells, prior to expression of mutant Ser-tRNAs, are already deficient at decoding of CUC and AGG codons and are under proteotoxic stress due to elimination of their cognate tRNAs (64), so the accurate decoding of the highlighted genes related to PKA metabolism is already compromised. This may indicate that destabilization of the endogenous tRNA pool may induce further defects in metabolism and/or cell cycle progression, and that the molecular mechanisms employed by the cell to counteract such destabilizations are energetic and time costly to the cell, as reflected by their growth rates.

Chapter 4

4. Evolution of codon ambiguities

4.1 Overview

The host laboratory has made a major breakthrough by engineering the first complete codon reassignment. Using tRNA engineering and gene deletion, they constructed a series of *C. albicans* strains that misincorporate Leu at the atypical Ser-CUG sites, at levels that range from 20% up to 99%. Surprisingly, these strains adapted to increasing Leu misincorporation, recovered growth rate to wild-type levels and displayed unanticipated phenotypic variability, including highly variable colony and cell morphologies, and increased tolerance to antifungals (58). As suggested by this and other studies (57, 70, 102), proteome destabilization produced by codon ambiguity is at some degree tolerable, and can be advantageous in some circumstances. So it has been suggested that codon ambiguity provides a mechanism to codon evolution, as far as the universe of statistical proteins produced provides phenotypic variability and innovation that allow organisms to explore new environments (48, 58, 103). Additionally, genome re-sequencing of the reassigned *C. albicans,* engineered by Bezerra and colleagues, showed rapid accumulation of unique single nucleotide polymorphisms (SNPs) and loss of heterozygosity (LOH) with increasing CUG ambiguity (58). These data show an intrinsic link between the genome and the genetic code, and support the hypothesis that codon reassignments can be engineered using forced evolution, releasing the code its frozen state.

Here, we will test the hypothesis that codon reassignments can be engineered using forced evolution, by subjecting the ambiguous strains described in chapter 3 to experimental evolution. These strategies allow the maintenance of yeast populations over timescales spanning many hundreds or even thousands of generations, in which changes in both genotype and phenotype can be monitored. There are three main strategies to propagate organisms in experimental evolution studies, which explore different types of genetic dynamics (figure 4.1) (74). In mutation accumulation experiments (figure 4.1a), the effects and rates of a new mutations are studied by imposing population bottlenecks. These minimize the size of the population which reduces genetic diversity and may lead to the fixation of arbitrary mutations (75). In adaptive evolution experiments, the genetic dynamics are driven by preferential accumulation of mutations that are better adapted to certain environments. Our evolutionary experiment is part of the last group and

was made by serial transfer (figure 4.1c), in which a part of the yeast population is periodically transferred to fresh media and allowed to regrow until the limiting nutrient is exhausted. By doing so, the genetic diversity is maintained at each transfer (104).

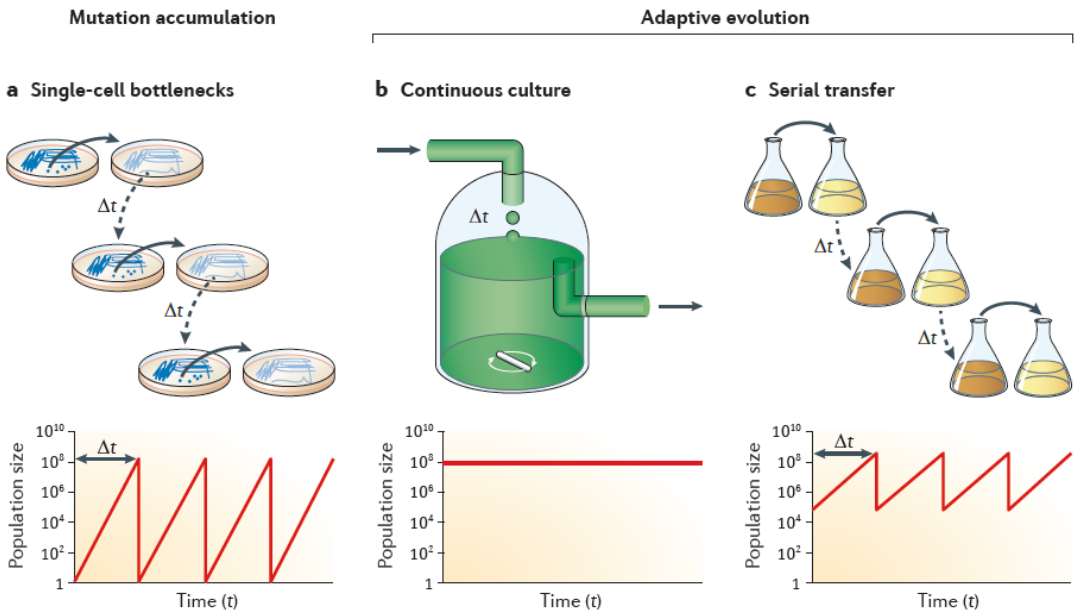


Figure 4.1 – Different types of experimental evolution. Adapted from Barrick *et al.*, 2013. (a) Mutation accumulation experiments. (b and c) Adaptive evolution by (b) continuous cultures and by (c) serial transfer.

4.2 Results

4.2.1 Evolution of the Leucine set

Cells were subjected to experimental evolution, until they reached ~200 generations. After this point was reached (around 20 passages, every 2-3 days depending on the duration of the lag phase) the fitness parameters were accessed again. As before, low and high expressing groups had different behaviors.

After evolution, low expressing strains (SC) showed a similar trend to the control strains ΔtL(GAG)G and ΔtL+EV (figure 4.2), that is a small decrease in growth rate when compared to their non-evolved counterpart. ΔtL(GAG)G had a 4% decrease, while ΔtL+SC-tRNA(GAG) showed a ~14% decrease in growth rate after evolution (paired t-test, P=0.008). ΔtL+EV and ΔtL+SC-tRNA(GAG) had 25% and 19% decrease in growth rate, respectively. This behavior is completely different from the observed in the high-expressing group (MC), where an increase in growth rate was observed in comparison to their non-evolved counterpart. This shows that cells coped differently with different levels of the same tRNA.

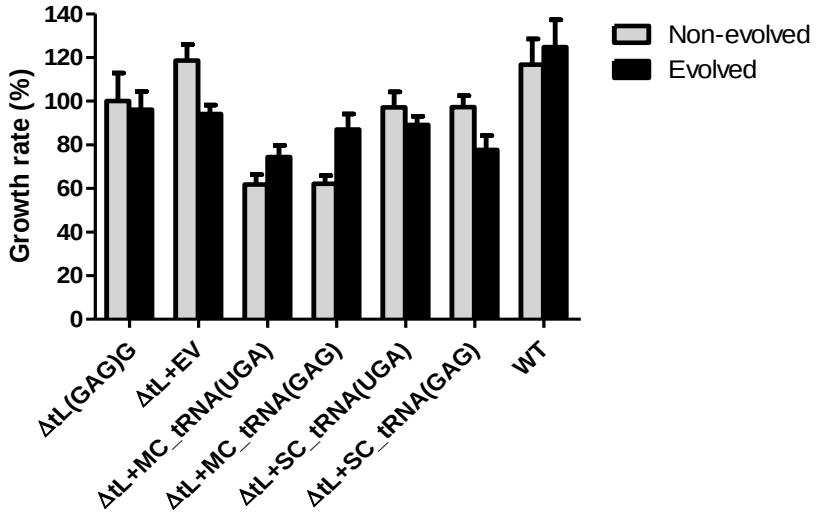


Figure 4.2 – Comparison between the growth rate of non-evolved (grey) and evolved (black) Leucine mutants. Data represents the mean + SD of 5 different clones.

Strains expressing both Ser-tRNA$_{UGA}$ and Ser-tRNA$_{GAG}$ at high level, recovered significantly in several growth parameters after 200 generations. Indeed, EΔtL+MC-tRNA(UGA) and EΔtL+MC-tRNA(GAG) recovered approximately 15% and 25% of their growth, respectively (paired t-test, *P*=0.03 and *P*=0.0014, respectively). In this way, cells expressing the mutant tRNA at high level reached growth rate levels similar to the observed in the deletion strain ΔtL(GAG)G. Although these strains did not achieve WT growth rate levels, this astoneshing recovery led us to hypothesize that it may take more time to fully achieve the adapted state. So it would be interesting to evaluate their behavior after another round of evolution. Our thought is that these ambiguous strains would accumulate a series of beneficial mutations that would adapt translation to the new genetic background, as in the work of Bezerra and coworkers (58).

Another surprising recovery seen in these MC strains is in the duration of the lag phase (figure 4.3). Particularly, clones overexpressing the endogenous Ser-tRNA$_{UGA}$ (ΔtL+MC_tRNA(UGA)) took less than half the time required in the beginning of the experiment to enter exponential. Although less pronounced, cells highly expressing the mutant Ser-tRNA$_{GAG}$ also exhibited a recovery in this parameter, cutting almost 10 hours in the duration of lag phase. Together with their recovery in growth rate, this shows that these strains can adapt to the induced mistranslation. Similar results relative to this phase have been obtained after evolution of *E. coli* with extended lag phases. Also, accordingly with the reported in *E. coli* (105), these high-expressing strains showed a detectable recovery on the duration of this phase, as early as after 5 passages (105). Unfortunately, we only have data at mid-point evolution from the strain expressing Ser-tRNA$_{UGA}$, with ~120 generations,

and at this point they already had cut half of time required to enter exponential phase. As seen in bacteria, adaptations that reduce lag phase are highly beneficial, and are readily fixed before adaptations that benefit growth rate (105). However, the behavior of growth rate levels of our mutant strains were heterogeneous across clones (figure 4.4). Indeed, at half-time experiment (100G), each clone showed recovery rates ranging increases from 10% to 30%, relative to the non-evolved strain (blue bar in figure 4.4). Interestingly, the one clone that showed the highest growth recovery (31%) at generation 100 (clone 2), then slightly decreased the recovery (~8% compared to 100G) in the end of the evolution. The other clones had a more moderate recovery at half-time evolution, and interestingly the clone who recovered less in the first phase (clone 1), was the one showing the better performance at the end of the experiment (with an overall increase of 32% relative to non-evolved clone, and around 20% recovery in the second phase).

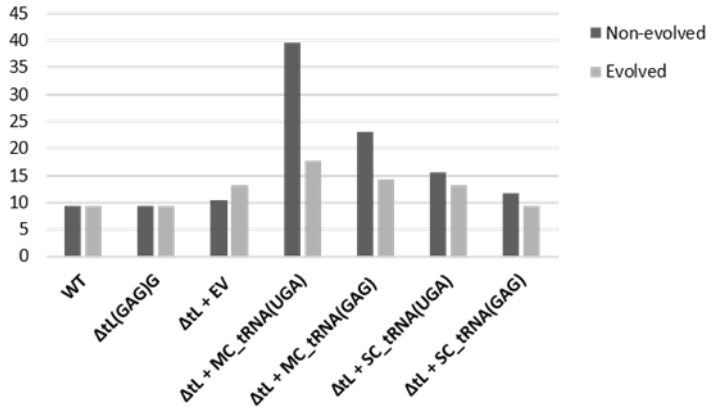


Figure 4.3 – Comparison between the lag phase duration, in hours, of non-evolved (dark grey) and evolved Leucine mutants (light grey). Data represents the average of 5 different clones.

Another difference between the two multi-copy strains is how clones from the two strains cope with mistranslation. As seen in figure 4.5, clones highly expressing the mutant tRNA after evolution (EΔtL+MC-tRNA(GAG)) constitute a more homogeneous population than clones overexpressing the endogenous $tRNA_{UGA}$. This subtle feature may indicate that different biological mechanisms may be at hands. Half of the clones overexpressing the endogenous Ser-tRNA had a better performance (recovered ~25% of their growth rate) and reached a higher growth yield (final OD_{600}, which indicates the amount of biomass produced) than the other half, which only recovered ~15% of their growth rate (figure 4.5B). On the other hand, clones with high expressing of the misreading tRNA are a homogeneous population (figure 4.5A), which may indicate that the same mechanisms and mutations are involved in adaptation to mistranslation in all clones.

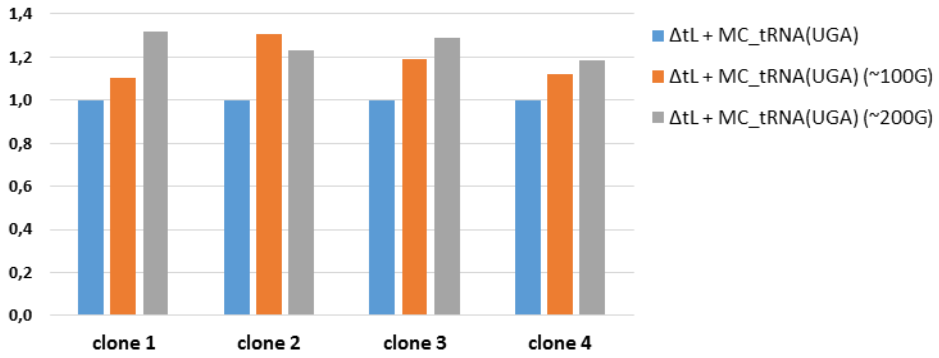


Figure 4.4 – Comparison between the growth rate of non-evolved ΔtL strains highly expressing the endogenous Ser-tRNA$_{UGA}$ (blue bar), and evolved strains with 100 (orange bar) and 200 generations (grey bar). Each clone is normalized to its non-evolved form.

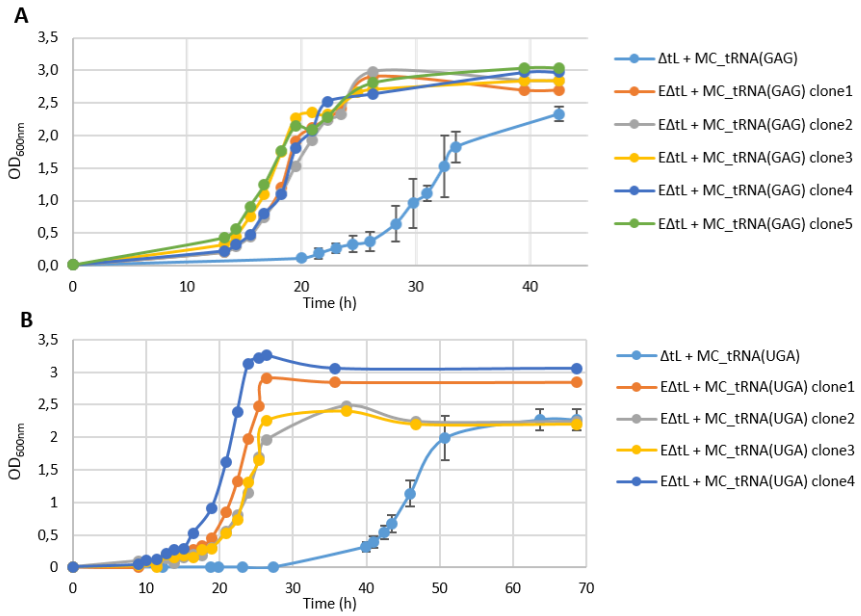


Figure 4.5 – Comparison between the growth curve profiles of non-evolved (ΔtL, light blue) and evolved (EΔtL) Leu mutants. Growth curve measurements as optical density (OD) units over time of mutants highly expressing Ser-tRNA$_{GAG}$ (A) and Ser-tRNA$_{UGA}$ (B). Non-evolved data represents the average ± SD of 5 different clones (light blue line)

4.2.2 Evolution of the Arginine set

Contradictory to the work done by Yona and colleagues (106), we did not see the recovery of the deletion strain ΔtR(CCU) after evolution to WT levels. Indeed, here, evolved ΔtR(CCU) (EΔtR) shows a slight decrease and a homogenization of the population (figure 4.6). Again, strains expressing tRNAs at low level (SC) did not exhibit meaningful growth phenotypes. The only

observation of note, is that beside the decrease seen in growth rate of ΔtR+SC-tRNA(CCU), the population became more homogeneous and intriguingly fixated in lower growth phenotype.

As in the Leucine set, strains expressing the tRNAs at high levels (MC) showed a significant recovery in growth rate (paired t-test, ΔtR+MC-tRNA(UGA) vs. EΔtR+MC-tRNA(UGA): *P*=0.027; and ΔtR+MC-tRNA(CCU) vs. EΔtR+MC-tRNA(CCU): *P*=0.043). However, the pattern of alteration in the duration of the lag phase (figure 4.7) is not like the one of the Leu set, highlighting the development of different metabolic and genetic networks in the Arg set. For example, a different mechanism may be at play in cells expressing Ser-tRNA$_{CCU}$ at high level, as they recover their growth rate without altering the time required for entering exponential phase. This is also contrary to the *E. coli* study, where lag phase adaptations are always fixed before growth rate improvements (105).

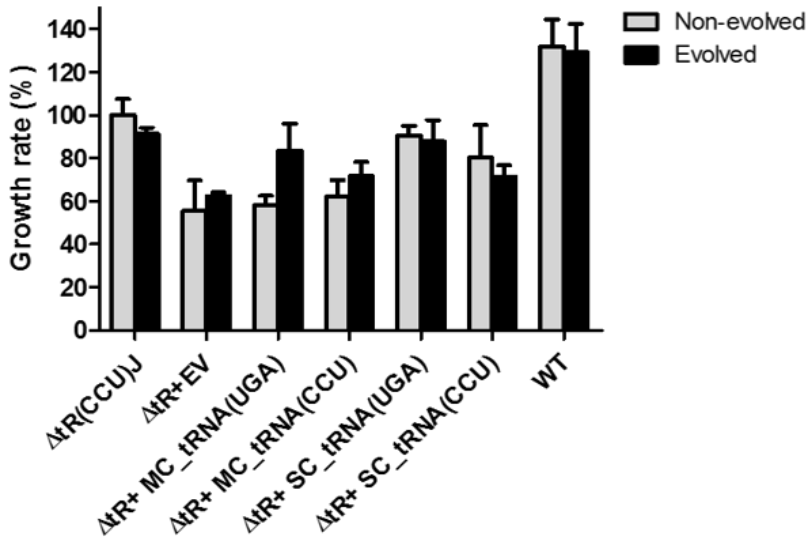


Figure 4.6 – Comparison between the growth rate of non-evolved (grey) and evolved (black) Arginine (ΔtR) mutants. Data represents the mean + SD of 5 different clones.

Cells overexpressing the endogenous tRNA were the ones that showed the highest recovery in both parameters (figures 4.6 and 4.7). Still, like in the Leucine set, the population was heterogeneous and could be divided in two sub-sets, according to growth rate: one that recovered ~25% of their initial growth rate, and other that surprisingly recovered ~60% (data not shown). So it seems that in the genetic background of ΔtR(CCU), destabilizations in the endogenous tRNA pool are better tolerated than substitutions of Arg to Ser in response to codons AGG, unlike the Leucine set.

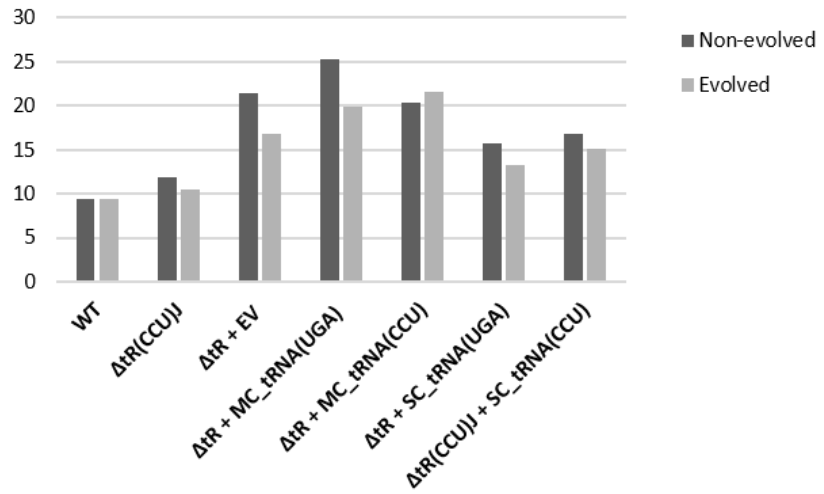


Figure 4.7 – Comparison between the lag phase duration, in hours, of non-evolved (dark grey) and evolved (light grey) Arginine mutants. Data represents the average of 5 different clones.

4.2.3 Northern blot analysis of mutant tRNA expression

Several mechanisms may be behind the recovery in fitness of the evolved mistranslating strains. Robustness may emerge from various processes, from transcription level to changes in the chaperone system. Other hypothesis is the mitigation of the error rate during the evolution by decreasing expression of the mutant tRNA. To verify if the growth recovery observed in the multi-copy expressing strains could be due to a decrease in the chimeric tRNA expression, a northern blot analysis was performed. All tRNAs were detected, which shows that in fact they are being expressed and the observed fitness phenotypes, both before and after evolution, are attributable to the mutant tRNA (figure 4.8).

Interestingly, expression of Ser-tRNA$_{GAG}$ appears to increase after evolution. Indeed, quantification by ImageJ software showed an increase in Ser-tRNA expression in all evolved strains. The huge increase seen in the evolved strains expressing Ser-tRNA$_{UGA}$ and Ser-tRNA$_{GAG}$ (clone3) may be due to bad quality sample, as blots appear as smears. The other two evolved clones with high expression of the mutant Ser-tRNA$_{GAG}$ (EΔtL+Ser-tRNA$_{GAG}$) showed increases of 30% and 78% in expression. This may indicate that these cells have adapted to mistranslation, and somehow have selected it as beneficial during evolution. Together with the previous results, the hypothesis of a codon reassignment, at least at some degree, seems more promising.

The Arginine set behaved differently in the tRNA expression parameter too. To alleviate the deleterious impact of mistranslation, the trend revealed by this set seems to culminate in lowering the expression of the mistranslating tRNA. Consistently, evolved mutants expressing Ser-tRNA$_{CCU}$

showed a decrease of around 40% in tRNA expression, while the tRNA disrupting the endogenous tRNA pool showed a more modest decrease of 20%. Again, striking differences arose from the two sets, indicating that different cellular and molecular mechanism may be at action.

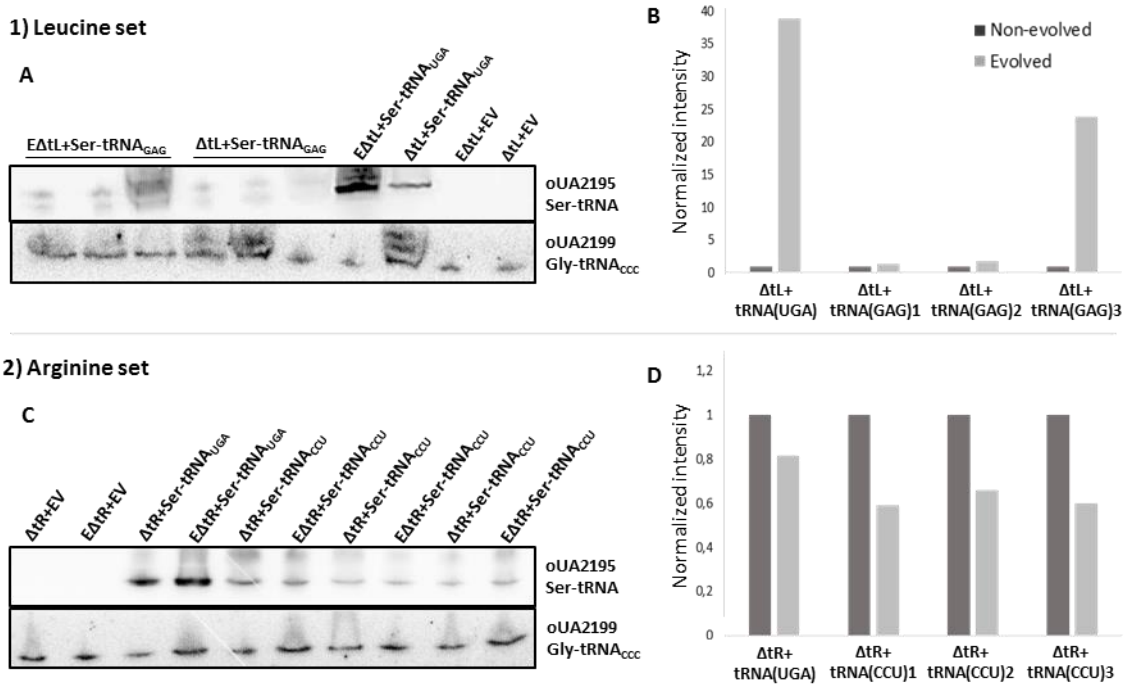


Figure 4.8 – Northern blot analysis of the mutant tRNAs. Northern blot hybridizations were carried out using ɣ-^{32}P-ATP-tRNA$_{UGA}^{Ser}$ (oUA2199) and ɣ-^{32}P-ATP-tRNA$_{CCC}^{Gly}$ (oUA2195) probes. Panel 1 shows the hybridization membranes from the Leu set (A) and their intensity quantification (B). Panel 2 shows the hybridization membranes from the Arg set (C) and their intensity quantification (D). Quantification of each pair of un-evolved (dark grey) and evolved clones (light grey) were normalized to the respective non-evolved version (=1)

4.2.4 Mutant tRNA sequencing

To monitor if the inserted tRNAs were mutated or degraded, and so validate these experiments, we extracted the plasmids containing the engineered tRNAs from non-evolved and evolved yeast mutants, and sequenced them. Unfortunately, strains from the arginine set did not exhibit the altered anticodon, which invalidates data relative to the mutant Ser-tRNA$_{CCU}$ set. Since the constructed plasmids were sequenced and validated prior to cloning experiments, so one out of two possibilities exist: or strains unexpectedly reverted the codon identity of engineered tRNAs, or it could be a simple human error in the begging of the experiment or during plasmid extraction. Still, data from strains overexpressing Ser-tRNA$_{UGA}$ are still valid. Indeed, beside the disappointing

sequencing results, strains expressing Ser-tRNA$_{UGA}$ and Ser-tRNA$_{CCU}$ showed different behaviours consistently. But still we decided to not use this set in further analysis.

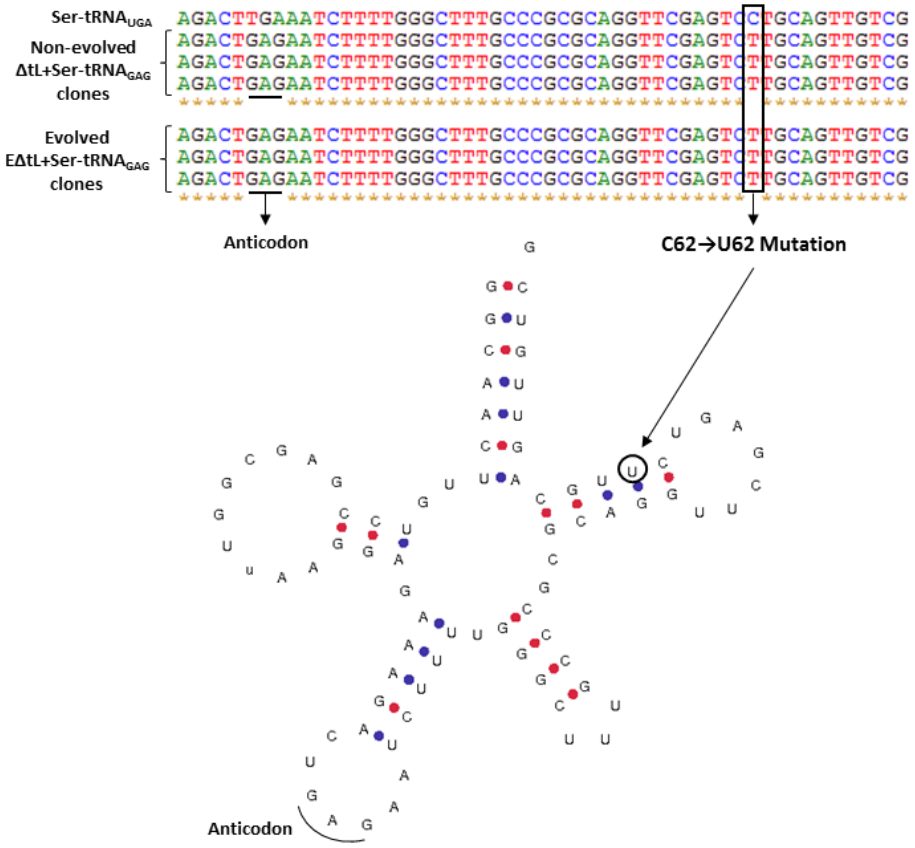


Figure 4.9 – Sequencing data from strains highly expressing Ser-tRNA$_{GAG}$ and the theoretical structure of the mutant tRNA. First line of the sequencing data is relative to the WT Ser-tRNA$_{UGA}$, while the others show a representative sample of non-evolved and evolved clones highly expressing Ser-tRNA$_{GAG}$. C to U mutation and the anticodon are highlighted both in the sequencing data and in the tRNA structure.

The totality of the Leucine set was validated by sequencing. Interestingly, we found that all the non-evolved and evolved clones expressing Ser-tRNA$_{GAG}$ in a multi-copy plasmid acquired a mutation C→U in position 62, which is located in the TψC arm (figure 4.9). Yap and colleagues mutagenized this position to other nucleotides in *E. coli* Pro-tRNA and concluded that as long as wobble pair is maintained, like G_{52}-U_{62} seen in our mutant tRNA, no major defect is seen in the tRNA aminoacylation (107). Singh and Green also reported that the C62U mutant of Met-tRNA$_i$ had no defect in nuclear export (108). The only functional implication found across the bibliography was its influence in modifications in position 55 of the T-loop, and at a less extent in position 54. Becker and coworkers constructed a series of mutant T-loop mini-helixes as substrate for the modification enzymes m^5U_{54}-methyltransferase and ψ_{55}-sintetase, and found that the introduction of a wobble

pair G_{52}-U_{62}, in place of a Watson-crick pair G_{52}-C_{62}, dramatically decreased (but did not abolished) the efficiency of ψ_{55} formation in yeast Asp-tRNA$_{GUC}$ (109). Pseudouridine-55 is found in almost all tRNA species from yeast (110) (except in the initiator Met-tRNA (111)) and contributes to the stabilization of specific structural motifs in the TψC-loop (111, 112). Noteworthy, the disruption of the yeast gene *PUS4*, encoding the sole enzyme capable of ψ_{55} synthesis, has no effect on cell growth. Given its almost universal conservation, this is somehow unexpected, but previous studies had already shown that the absence of ψ_{55} does not impact tRNA aminoacylation, nor it affects translation *in vitro* (113). So the functional consequences of the mutation found in all the clones remains a mystery, but its degree of homogeneity implies that they are in fact functional, and may be linked to fine-tuning of specific steps in protein synthesis (113).

4.2.5 Monitoring of Ser misincorporation

In order to monitor Ser misincorporation by at CUC-Leu and AGG-codon in cells with high expression of mutant tRNAs, we engineered a reporter system based on the yeast-enhanced green fluorescent protein gene (*y*EGFP) (114).

Before we knew the sequencing results from the Arginine set, we had already constructed the reporter system for this set (figure 4.10). Although the results obtained from the strains expressing the mutant tRNA are not valid, this method seems suitable to this kind of study and can be used in future studies. Position 96 of GFP has been reported as critical to proper chromophore formation, and only exhibit fluorescence if Arg or Lys are incorporated at this position. Since any other amino acid substitutions produces inactive and non-fluorescent proteins, the authors rationalized that the mutant proteins did not properly fold or were poorly expressed (115). So the manipulation of the codon encoding this residue seemed perfect to monitor misincorporation of Ser residues instead of the Arg ones. Indeed, the mutagenesis of the AGA-Arg codons to UCU-Ser codons rendered cells with no fluorescence, contrary to cells expressing the wild-type version of GFP, where these codons are decoded as Arg.

In the case of the Leu set (figure 4.11), we engineered position 201 of GFP, which is encoded by a UUA-Leu codon, as it was described as critical to protein stability and any substitution renders GFP inactive (114). We used a similar strategy reported by the host laboratory (58), but in this case, the reporter system is a loss-of-function one. For this, we mutated the codon encoding residue 201 to the “orphan” CUC-Leu codon. This GFP version serves as a reporter, because if Ser is inserted by the mutant Ser-tRNA$_{GAG}$ in the place of Leu, cells would have no fluorescence, allowing us to

monitor its misincorporation. As a negative control, we constructed a UCU codon at this position, which is always decoded as Ser, leading to the quick GFP degradation (58).

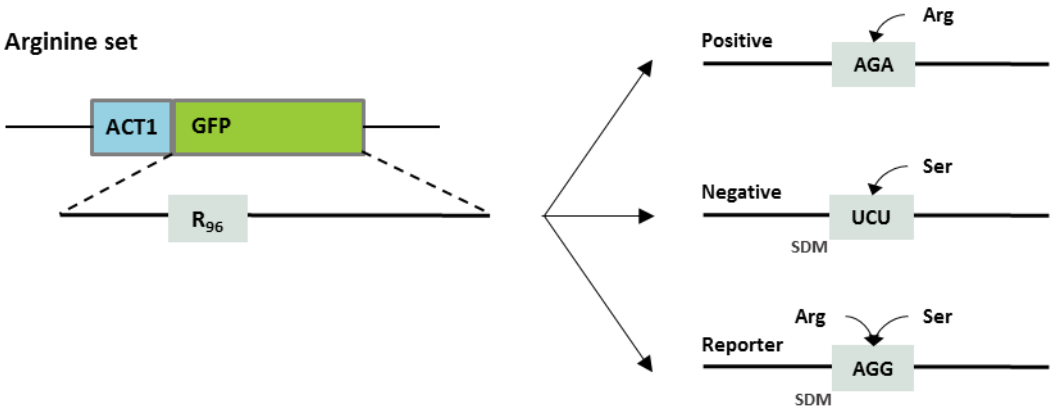


Figure 4.10 – Scheme of the reporter system built to monitor serine incorporation at AGG-Arg codons. The system is based on the plasmid pACT1-GFP, which contains the codon-optimized yeast enhanced green fluorescent protein (*y*EGFP) (47). The arginine-96 is encoded by the AGA-Arg codon which was mutated to UCU-Ser (negative control) and to the "orphan" AGA-Arg codon (reporter).

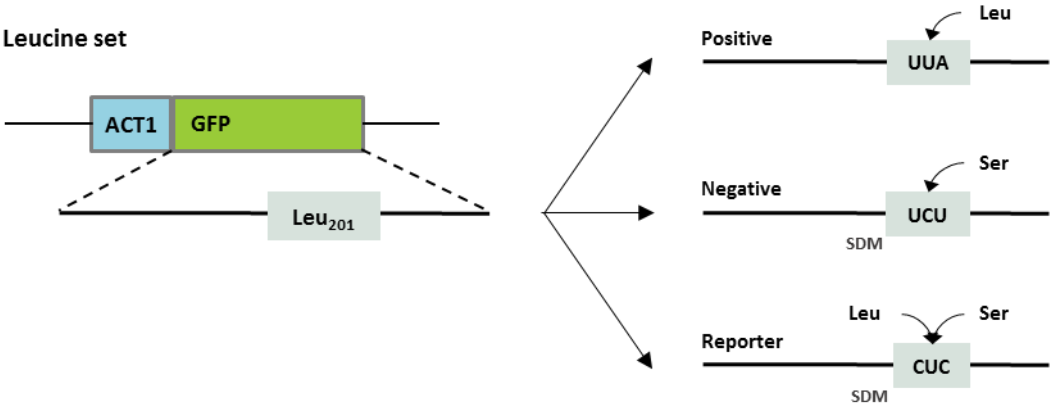


Figure 4.11 – Scheme of the reporter system built to monitor serine incorporation at CUC-Leu codons. The system is based on the plasmid pACT1-GFP, which contains the codon-optimized yeast enhanced green fluorescent protein (*y*EGFP) (47). The leucine-201 is encoded by the UUA-Leu codon which was mutated to UCU-Ser (negative control) and to the "orphan" CUC-Leu codon (reporter).

In order to express the plasmids harboring the GFP constructions in evolved cells, we first had to remove the plasmid containing the mutant tRNA, as both vectors carry the URA3 marker. Following the notion that toxic plasmids are promptly discarded by cells grown in media without selection, we first grew them in YPD and after a few passages, colonies where again stroked in the selectable media. Surprisingly, cells did not lose plasmid across 20 passages (corresponding to almost ~200 generations, another round of evolution). This is indicative that evolved strains (and, at a lesser extent, un-evolved strains) are strongly selecting the mutant tRNA across generations,

even when not exposed to selective conditions. Consequently, we then used a more efficient technique. Cells were then plated in 5-FOA plates, which only enables growth of *ura3*$^{-}$, since it is converted by the enzyme encoded by URA3 to fluorodeoxyuridine, which is toxic to cells (116).

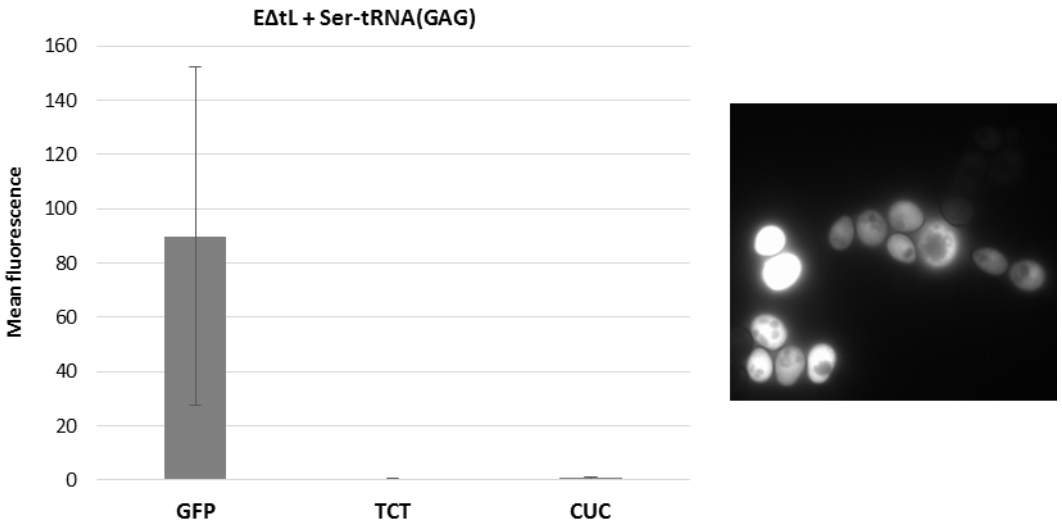


Figure 4.12 – Quantification of the mean fluorescence of evolved Leu mutants expressing Ser-tRNA$_{GAG}$. The right picture is a representative photo of positive clones, showing the population heterogeneity**.**

Surprisingly, evolved strains expressing the mutant Ser-tRNA$_{GAG}$ exhibit the same behavior as the negative control, showing a complete loss of GFP function (figure 4.12). This indicates that mutant tRNAs are indeed incorporating Ser at CUC-Leu codons. In fact, these results imply that Ser miscorporation increased nearly to 100%, which indicates that this strains may have possibly achieved codon reassignment. However, these exciting results must be taken carefully, as this reporter system is not perfect, and is merely indicative. As one can see in the representative picture on figure 4.12, fluorescence is very heterogeneous and its signal is not homogeneous across the cell cytoplasm (for example, GFP does not seem to enter vacuoles). So, isogenic cells in one population show a broad range of fluorescence, indicating that this system is highly unstable. A good approach to this problem would be by expressing *y*EGFP using an integrating strategy, like previously used (58). Unfortunetly, transformations of non-evolved and evolved knock-out strains with the complete set of control plasmids (harboring only the different versions of GFP) proved recurrently unsuccessful. So apart from the innerent frailitys of this system, one should take this consideration into account.

Since quantification of the reporter system involves microscopic work, we used the acquired pictures to examine other morphological parameters, like area. Strains expressing Ser-tRNA$_{UGA}$ and

Ser-tRNA$_{GAG}$ were bigger after evolution, a phenotype that was not seen in the evolved strain with the empty plasmid (figure 4.13). This may indicate that evolved strains developed aneuploidies. Studies haves shown that aneuploidy induces genomic instability and mutagenesis in yeast (117). On the other hand, aneuploidy has also been linked to phenotypic variation and growth advantage in some conditions, which raises the hypothesis that in adverse environments, aneuploidy can provide a mechanism for cells to frisk phenotypic variations and thus adapt (118).

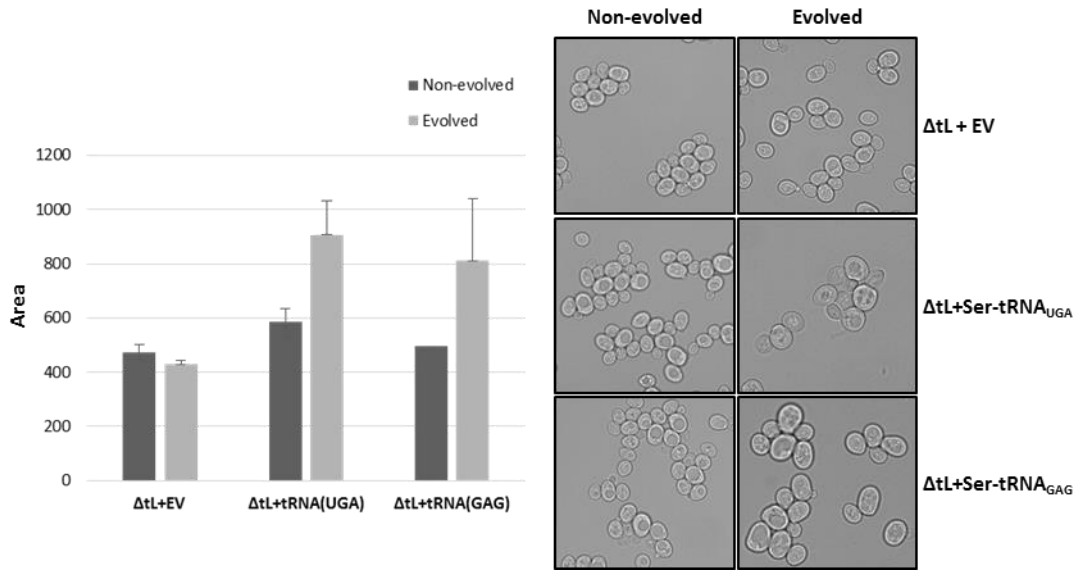


Figure 4.13 – Area measurements from non-evolved and evolved Leu mutants. The panel at the right has representative pictures (at scale) from each group.

Beside the dubious results from the Arginine set, un-evolved and evolved mutants were also visualized under the microscope. Although none of the later parameters were quantified, the morphology of these cells highlight again the differences between the two sets, as can be seen in figure 4.14. Regarding the un-evolved forms, cells were not different from cells belonging to the Leucine set. Only a slight tendency to more elliptical and aggregated forms was seen in mutants with high expression of the mutant Ser-tRNA$_{CCU}$. After 200 generations, the later presented abnormal cells with bigger, elongated and lobulated forms that tend to aggregate. Interestingly, cells evolved with low expression of the mutant Ser-tRNA$_{CCU}$ and later transformed with the reporter plasmid (note that the reporter backbone is a multi-copy vector) consistently show a flocculated phenotype. Together, this data show that these clones are under a higher stress than the Leucine ones and that the underlying mechanisms involved in each sets are different. Worthy of note, is the visual differences between cells expressing Ser-tRNA$_{UGA}$ and Ser-tRNA$_{CCU}$ that

together with the reported differences in growth rate, lag phase duration and tRNA expression seems to indicate that they are not isogenic as sequencing data suggests.

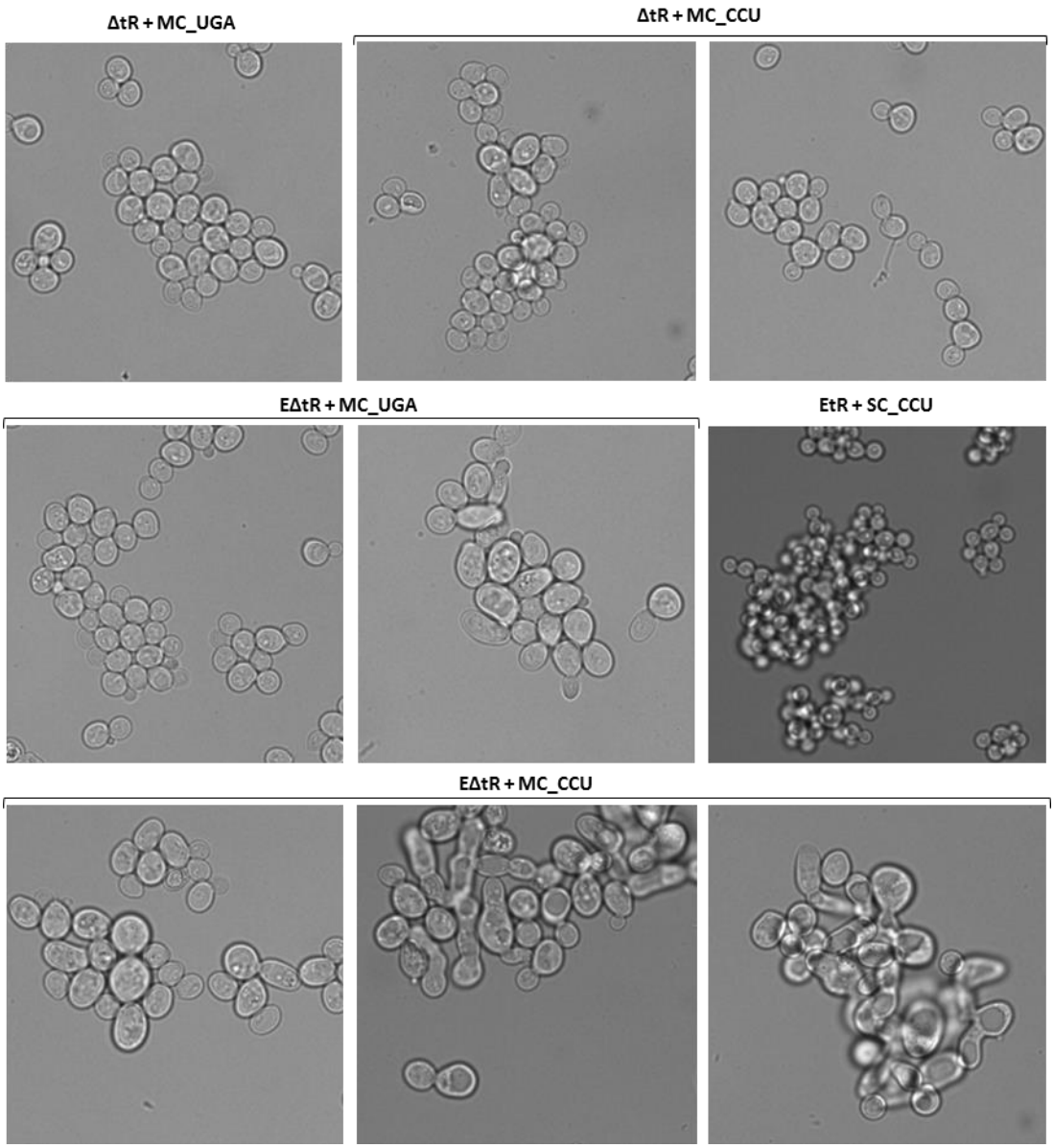


Figure 4.14 – Representative pictures of non-evolved and evolved Arg mutants. The upper row shows un-evolved strains with high expression of Ser-tRNA$_{UGA}$ (ΔtR+MC_UGA) and Ser-tRNA$_{CCU}$ (ΔtR+MC_CCU); middle row shows evolved strains with high expression of Ser-tRNA$_{UGA}$ (EΔtR+MC_UGA) and strains evolved in the low copy version of Ser-tRNA$_{CCU}$ (EΔtR+SC_CCU); the last row shows pictures from evolved strains with high expression of Ser-tRNA$_{CCU}$ (EΔtR+MC_CCU)

4.2.6 Discussion

After evolution, the strong negative effects produced by high expression of mutant and endogenous tRNAs were minimized, showing that cells have undergone a process of adaptation.

Particularly, Leu mutants recovered significantly both in growth rate and in duration of lag phase. Also, the fluorescent reporter system constructed to monitor Ser misincorporation at CUC-Leu sites, showed that serine was indeed incorporated and possibly codon reassignment was achieved. These data shows that these strains adapted to mistranslation.

Indeed, all the observations relative to Leu mutants expressing Ser-tRNA$_{GAG}$ at high level are consistent with the exciting idea of a possible codon reassignment. Through the evolution experiment, cells kept expressing the mutant tRNA and even showed an increase in its expression at the end of the experiment. Also, evolved strains were able to go through another complete round of evolution in rich media without losing the plasmid. We cannot assertively speculate the functional implication of the mutation C62U found in all Ser-tRNA$_{GAG}$ clones, but one must assume, since there was a recovery at the end, that during the course of evolution, if a preferential mutation accumulates, so it must be to better capacitate the cell to their genetic environment. Together these results indicate that the tRNA is being selected as beneficial and it is selectively transmitted to the offspring, even when no pressure to maintain the plasmid is kept.

Also, after evolution, these cells apparently developed aneuploidies, which is indicative of genome instability. It has been purposed that codon ambiguities destabilize the proteome in a dynamic way producing fluctuating genome instabilities of relevance to evolution and adaptation. In fact, codon ambiguity and evolution induces major changes in genome structure and gene expression (58). The developed aneuploidy may possibly be due to this genetic dynamic, although we do not have enough information to make such a statement. Sequencing the whole genome of our recombinant strains would provide valuable data if the acquired tolerance to codon ambiguity is the result of the accumulation of compensatory mutations along the evolution cycles, like single nucleotide polymorphisms (SNPs), small insertions and deletions (indels), gene copy number variation (CNV), loss of heterozygosity (LOH) or karyotype alterations.

Chapter 5

5. Discussion

5.1 General discussion and main conclusions

In recent years, remarkable efforts have been made to integrate unnatural amino acids with novel chemical properties into proteins (74). However, the main focus of these studies was production and characterization of novel proteins rather than codon reassignment. In this thesis, we tackled the later problem by generation of ambiguity at specific codons by an experimental strategy which involved deletion of single-copy tRNA genes followed by expression of non-cognate tRNAs that are able to compensate the decoding of the deleted tRNAs. Since the targeted codons are "orphan" and decoded inefficiently via wobble interactions by other iso-acceptor tRNAs (64, 85, 87), our system is expected to be highly detrimental to the cell homeostasis but was viable and different lineages adapted differently to codon ambiguity during an experimental evolution procedure. This work provides a series of proof-of-principle experiments that confirm that our experimental approach is valid, since recombinant strains, even those with high level expression of misreading tRNAs, remained viable.

Data from online codon usage databases (119) shows that in the *S. cerevisiae* nuclear genome, CUC-Leu and AGG-Arg codons have a usage frequency of 5.5‰ and 9.2‰ respectively, which illustrates its low usages (96). However CUC codons represents 17499 and AGGs represents 28483 codons across the yeast ORFeome, so misincorporation of serine across all these sites is expected to cause major proteome disruption. Indeed the presented results in this work shows the harsh deleterious consequences induced by ambiguous decoding. Cells with high expression of Ser-tRNAs showed extended lag phases and their growth rates dropped around 50% (figures 3.4 and 3.5), in relation to WT and deletion controls, which is concordant with high synthesis of aberrant proteins and indicates also that ambiguous strains may have defects in cell cycle progression or metabolism.

Indeed, bioinformatics codon content analysis suggests that cAMP related pathways may be compromised in our cells, as a series of key components of the PKA metabolic pathway have high frequency of these codons (figure 3.7). Since our mistranslating system has the potential to incorporate one of two amino acids at each one of these codons (Ser or Leu in ΔtL, and Ser or Arg in ΔtR), so translation of these genes will produce a group of statistically related proteins, and some of them will have deficient folding and function, as errors accumulate in their sequences.

Particularly, the gene that encodes the adenylate cyclase responsible for cAMP synthesis (CYR1) has 16 CUCs and 25 AGGs, so it is expected that our strains have low intracellular cAMP concentrations, which will lead to low cAMP-dependent kinases activity, like PKA. Since inactivation of adenylate cyclase, PKA or Cdc25 results in similar phenotypes of permanent arrest at G_0, unless cAMP is provided to the cells (120, 121), one can assume that some of the statistical proteins produced by our mutants are indeed functional as cells are viable and, although the observed delay in growth, they are able to surpass the cell cycle arrest.

One of the most intriguing and unexpected phenotypes was observed upon major deregulations of the endogenous tRNA pool by high expression of the seryl $tRNA_{UGA}$. Elevated tRNA levels are reported in pathologies like ovarian, breast and cervical cancers, due to aberrant RNA polymerase III expression (122-124). Interestingly, it has been reported that overexpression of Met-$tRNA_i$ in non-tumorigenic human breast epithelial cell lines induced increased levels of the other tRNAs, which resulted in increased metabolic activity and proliferation (125). However, information about the consequences of increased tRNA concentration are scarce. In *E. coli*, overexpression of Leu-$tRNA_{CAG}$ led to very slow growth and inhibition of protein synthesis. The authors showed that elevated levels of tRNA led to titration of LeuRS as only 40% of tRNAs were aminoacylated, and of modifying enzymes as part of the tRNA were hypomodified. Also, they reported reduced rate of protein synthesis with negligible misreading and that particular proteins were over-produced (126). Another interesting relation arose from the suppression of several amber and ochre codons upon overexpression of Gln-tRNA with CUG (127) and UUG (128) anticodons, respectively. On the other hand, tRNA selection by the ribosome is affected by tRNA concentration and more abundant tRNAs are preferentially selected as they are more rapidly translated (3). So one can hypothesize that due to increased Ser-$tRNA_{UGA}$ expression, decoding of Ser-UCA codons is faster, which will produce an imbalance in translational speed, and increases in translational elongation speed are known to produce a trade-off of decreased accuracy (129, 130). Since UCA codons are rather frequent in the genome, with a usage frequency of 18.7‰ in yeast, this potential effect is extended to a large set of genes (ANACONDA codon content analysis revealed that only 8.6% of yeast genes do not have this codon) and may function as “disruption hotspots” of translation speed and accuracy. Importantly, one must always have in mind that our strains already have disturbed translation, due to the loss of the single cognate tRNA capable of decoding CUC or AGG codons. We cannot make an assertive guess based on the data presented, but we can speculate that any of the disturbances above have the potential to further aggravate fidelity of translation by further destabilizations of the endogenous tRNA pool. Our results, together with these observations shows how little we

know about how cells respond to perturbations in cellular tRNA levels and how they regulate the individual expression of each tRNA.

The harmful consequences of erroneous translation have been widely reported, however information about long-term evolutionary responses to mistranslation is not known. We explored the hypothesis that forced evolution, in combination with ambiguous decoding, has the potential to release codons from their frozen state by evolving ambiguous strains throughout 200 generations. The Leucine mutants mostly affected by seryl tRNA expression, namely those with high expression of Ser-tRNAs (MC), recovered significantly in growth rate and in the time need to enter exponential growth (figure 4.2 and 4.3), which indicates that our mutants adapted to mistranslation. Given the fact that *C. albicans* naturally decodes the ambiguous CUG-Leu codon as Ser in the cytoplasm, one can hypothesize that these type of ambiguous decoding is tolerable, beside the inherent negative of a proteome-wide substitution.

Interestingly, all Leucine mutants (un-evolved and evolved) with high expression of the misreading Ser-tRNA$_{GAG}$ presented a single mutation in the TψC arm, where C_{62} changed to U_{62}. Also, the observed phenotypes are not due to tRNA degradation, as northern blot analysis revealed that its expression even increases throughout evolution. Although the functional consequences of the observed mutation remains elusive, on must assume that they must be functional in some way, as it is present in every single clone from this group.

Consistent with our expectation, the reporter system built to monitor Ser misincorporation at Leu-CUC sites, revealed a complete loss of function in strains highly expressing Ser-tRNA$_{GAG}$, which indicates that high rates of Ser is being misincorporated at Leu-CUC codons. However, results taken from the fluorescent reporter experiment must be taken carefully, as they show signal heterogeneity and the results are solely indicative. Together, these results suggest that possible codon reassignment may have been achieved in these cells.

5.2 Future work

Yeast cells were able to tolerate high level misincorporation of Ser at rare and “orphan” Leu-CUC codons, which have lost its cognate tRNA, with high negative costs to cellular fitness. However, throughout evolution, cells were able to adapt to their mistranslating environment and recover 25% of their growth rate, achieving fitness levels similar to the ones seen in the deletion strain. The results obtained in this work raises a series of intriguing questions that are worth further investigation.

Although ambiguous decoding is highly detrimental due to synthesis of aberrant proteins, it has been shown to provide important selective advantages in *S. cerevisiae* under stress conditions, otherwise lethal (34). *Candida albicans* decodes naturally the CUG-Leu codons as Ser *in vivo*, and accommodates a basal level of Leu-mistranslation of ~3% at optimal conditions (70). By increasing its level of mistranslation novel phenotypes arise, which implies that codon ambiguity has the potential to expand the proteome (58, 70). So it would be very interesting to do a phenotypic screen over a large set of conditions, in order to clarify whether the ambiguous strains would produce adaptive phenotypic alterations.

Since codon ambiguity and evolution induces major changes in genome structure and gene expression (58), it would be interesting to perform whole-genome sequencing of our recombinant strains. This would provide insightful data to whether our ambiguous strains acquire tolerance to codon ambiguity through the accumulation of compensatory mutations along the evolution cycles, as fitness parameters seem to suggest.

In order to get a complementary insight on the mechanism of tolerance to codon ambiguity and reassignment, it would be of interest to perform a transcriptome analysis. Since codon ambiguities disrupt protein structure and activate the stress response (10) we expect high gene deregulation at the start of each evolution cycle and gradual attenuation along the evolution, as adaptive mutations accumulate in the genome.

The main frailty in this work is the reporter system used to assess ambiguity level that, as already discussed, is far from optimal. So an important aspect to be improved is the quantification of Ser misincorporation at the targeted codons, as for example by mass-spectrometry techniques. One possible approach is by using an assay previously developed in the host laboratory, which is based on a mass-spectrometry reporter construct that consists in a cassette with the codon of interest (CUC and AGG) inserted in the CaPGK gene. The codon-cassette is flanked by two thrombin cleavage sites that facilitate the purification of the short reporter peptide encoded by the cassette (131).

Another important aspect is to rethink the overall strategy for the Arginine set. As our results shows, the underlying genetic background of the ΔtR(CCU)J is complex and more sensitive to deregulations. To abrogate the impact of the insertion of the empty vector in these strains, we should rethink the selective marker used in the plasmid, since it appears that even one AGG codon in the URA3 has the potential to disrupt the enzyme encoded. However the auxotrophy of theses strains presents a limited set of alternatives and so this aspect must be carefully planned.

References

1. Line MA. The enigma of the origin of life and its timing. Microbiology. 2002;148(Pt 1):21-7.

2. Crick FH. The origin of the genetic code. Journal of molecular biology. 1968;38(3):367-79.

3. Higgs PG, Ran W. Coevolution of codon usage and tRNA genes leads to alternative stable states of biased codon usage. Molecular biology and evolution. 2008;25(11):2279-91.

4. Angellotti MC, Bhuiyan SB, Chen G, Wan XF. CodonO: codon usage bias analysis within and across genomes. Nucleic acids research. 2007;35(Web Server issue):W132-6.

5. Clancy S, Brown W. Translation: DNA to mRNA to Protein. Nature Education. 2008;1(1):101.

6. Schimmel P. Mistranslation and its control by tRNA synthetases. Philosophical transactions of the Royal Society of London Series B, Biological sciences. 2011;366(1580):2965-71.

7. Raina M, Ibba M. tRNAs as regulators of biological processes. Frontiers in genetics. 2014;5:171.

8. Rich A, RajBhandary UL. Transfer RNA: molecular structure, sequence, and properties. Annual review of biochemistry. 1976;45:805-60.

9. Ashraf SS, Ansari G, Guenther R, Sochacka E, Malkiewicz A, Agris PF. The uridine in "U-turn": contributions to tRNA-ribosomal binding. Rna. 1999;5(4):503-11.

10. Giege R, Juhling F, Putz J, Stadler P, Sauter C, Florentz C. Structure of transfer RNAs: similarity and variability. Wiley interdisciplinary reviews RNA. 2012;3(1):37-61.

11. Jackman JE, Alfonzo JD. Transfer RNA modifications: nature's combinatorial chemistry playground. Wiley interdisciplinary reviews RNA. 2013;4(1):35-48.

12. Agris PF. Decoding the genome: a modified view. Nucleic acids research. 2004;32(1):223-38.

13. Giege R, Sissler M, Florentz C. Universal rules and idiosyncratic features in tRNA identity. Nucleic acids research. 1998;26(22):5017-35.

14. Haseltine WA, Block R. Synthesis of guanosine tetra- and pentaphosphate requires the presence of a codon-specific, uncharged transfer ribonucleic acid in the acceptor site of ribosomes. Proceedings of the National Academy of Sciences of the United States of America. 1973;70(5):1564-8.

15. Ross W, Vrentas CE, Sanchez-Vazquez P, Gaal T, Gourse RL. The magic spot: a ppGpp binding site on E. coli RNA polymerase responsible for regulation of transcription initiation. Molecular cell. 2013;50(3):420-9.

16. Lloyd AJ, Gilbey AM, Blewett AM, De Pascale G, El Zoeiby A, Levesque RC, et al. Characterization of tRNA-dependent peptide bond formation by MurM in the synthesis of Streptococcus pneumoniae peptidoglycan. The Journal of biological chemistry. 2008;283(10):6402-17.

17. Roy H, Ibba M. RNA-dependent lipid remodeling by bacterial multiple peptide resistance factors. Proceedings of the National Academy of Sciences of the United States of America. 2008;105(12):4667-72.

18. Ling J, Reynolds N, Ibba M. Aminoacyl-tRNA synthesis and translational quality control. Annual review of microbiology. 2009;63:61-78.

19. Ibba M, Soll D. Aminoacyl-tRNA synthesis. Annual review of biochemistry. 2000;69:617-50.

20. Guo M, Schimmel P. Structural analyses clarify the complex control of mistranslation by tRNA synthetases. Current opinion in structural biology. 2012;22(1):119-26.

21. Kapp LD, Lorsch JR. The molecular mechanics of eukaryotic translation. Annual review of biochemistry. 2004;73:657-704.

22. Walsh D, Mohr I. Viral subversion of the host protein synthesis machinery. Nature reviews Microbiology. 2011;9(12):860-75.

23. Knight RD, Freeland SJ, Landweber LF. Rewiring the keyboard: evolvability of the genetic code. Nature reviews Genetics. 2001;2(1):49-58.

24. Ohama T, Inagaki Y, Bessho Y, Osawa S. Evolving genetic code. Proceedings of the Japan Academy Series B, Physical and biological sciences. 2008;84(2):58-74.

25. Meyer F, Schmidt HJ, Plumper E, Hasilik A, Mersmann G, Meyer HE, et al. UGA is translated as cysteine in pheromone 3 of Euplotes octocarinatus. Proceedings of the National Academy of Sciences of the United States of America. 1991;88(9):3758-61.

26. Lozupone CA, Knight RD, Landweber LF. The molecular basis of nuclear genetic code change in ciliates. Current biology : CB. 2001;11(2):65-74.

27. Keeling PJ, Doolittle WF. Widespread and ancient distribution of a noncanonical genetic code in diplomonads. Molecular biology and evolution. 1997;14(9):895-901.

28. Schneider SU, de Groot EJ. Sequences of two rbcS cDNA clones of Batophora oerstedii: structural and evolutionary considerations. Current genetics. 1991;20(1-2):173-5.

29. Bock A, Forchhammer K, Heider J, Leinfelder W, Sawers G, Veprek B, et al. Selenocysteine: the 21st amino acid. Molecular microbiology. 1991;5(3):515-20.

30. Srinivasan G, James CM, Krzycki JA. Pyrrolysine encoded by UAG in Archaea: charging of a UAG-decoding specialized tRNA. Science. 2002;296(5572):1459-62.

31. Santos MA, Tuite MF. The CUG codon is decoded in vivo as serine and not leucine in Candida albicans. Nucleic acids research. 1995;23(9):1481-6.

32. Sugita T, Nakase T. Non-universal usage of the leucine CUG codon and the molecular phylogeny of the genus Candida. Systematic and applied microbiology. 1999;22(1):79-86.

33. Soll D, RajBhandary UL. The genetic code - thawing the 'frozen accident'. Journal of biosciences. 2006;31(4):459-63.

34. Santos MA, Moura G, Massey SE, Tuite MF. Driving change: the evolution of alternative genetic codes. Trends in genetics : TIG. 2004;20(2):95-102.

35. Osawa S, Jukes TH. Codon reassignment (codon capture) in evolution. Journal of molecular evolution. 1989;28(4):271-8.

36. Miranda I, Silva R, Santos MA. Evolution of the genetic code in yeasts. Yeast. 2006;23(3):203-13.

37. Schultz DW, Yarus M. Transfer RNA mutation and the malleability of the genetic code. Journal of molecular biology. 1994;235(5):1377-80.

38. Lovett PS, Ambulos NP, Jr., Mulbry W, Noguchi N, Rogers EJ. UGA can be decoded as tryptophan at low efficiency in Bacillus subtilis. Journal of bacteriology. 1991;173(5):1810-2.

39. Ambrogelly A, Palioura S, Soll D. Natural expansion of the genetic code. Nature chemical biology. 2007;3(1):29-35.

40. Su D, Lieberman A, Lang BF, Simonovic M, Soll D, Ling J. An unusual tRNAThr derived from tRNAHis reassigns in yeast mitochondria the CUN codons to threonine. Nucleic acids research. 2011;39(11):4866-74.

41. Suzuki T, Ueda T, Watanabe K. The 'polysemous' codon--a codon with multiple amino acid assignment caused by dual specificity of tRNA identity. The EMBO journal. 1997;16(5):1122-34.

42. Schultz DW, Yarus M. On malleability in the genetic code. Journal of molecular evolution. 1996;42(5):597-601.

43. Tomita K, Ueda T, Watanabe K. 7-Methylguanosine at the anticodon wobble position of squid mitochondrial tRNA(Ser)GCU: molecular basis for assignment of AGA/AGG codons as serine in invertebrate mitochondria. Biochimica et biophysica acta. 1998;1399(1):78-82.

44. Lee JW, Beebe K, Nangle LA, Jang J, Longo-Guess CM, Cook SA, et al. Editing-defective tRNA synthetase causes protein misfolding and neurodegeneration. Nature. 2006;443(7107):50-5.

45. Zhao L, Longo-Guess C, Harris BS, Lee JW, Ackerman SL. Protein accumulation and neurodegeneration in the woozy mutant mouse is caused by disruption of SIL1, a cochaperone of BiP. Nature genetics. 2005;37(9):974-9.

46. Drummond DA, Wilke CO. The evolutionary consequences of erroneous protein synthesis. Nature reviews Genetics. 2009;10(10):715-24.

47. Chen B, Retzlaff M, Roos T, Frydman J. Cellular strategies of protein quality control. Cold Spring Harbor perspectives in biology. 2011;3(8):a004374.

48. Moura GR, Carreto LC, Santos MA. Genetic code ambiguity: an unexpected source of proteome innovation and phenotypic diversity. Current opinion in microbiology. 2009;12(6):631-7.

49. Allmang C, Wurth L, Krol A. The selenium to selenoprotein pathway in eukaryotes: more molecular partners than anticipated. Biochimica et biophysica acta. 2009;1790(11):1415-23.

50. Turanov AA, Lobanov AV, Fomenko DE, Morrison HG, Sogin ML, Klobutcher LA, et al. Genetic code supports targeted insertion of two amino acids by one codon. Science. 2009;323(5911):259-61.

51. Gaston MA, Jiang R, Krzycki JA. Functional context, biosynthesis, and genetic encoding of pyrrolysine. Current opinion in microbiology. 2011;14(3):342-9.

52. Netzer N, Goodenbour JM, David A, Dittmar KA, Jones RB, Schneider JR, et al. Innate immune and chemically triggered oxidative stress modifies translational fidelity. Nature. 2009;462(7272):522-6.

53. Luo S, Levine RL. Methionine in proteins defends against oxidative stress. FASEB journal : official publication of the Federation of American Societies for Experimental Biology. 2009;23(2):464-72.

54. Wiltrout E, Goodenbour JM, Frechin M, Pan T. Misacylation of tRNA with methionine in Saccharomyces cerevisiae. Nucleic acids research. 2012;40(20):10494-506.

55. Jones TE, Alexander RW, Pan T. Misacylation of specific nonmethionyl tRNAs by a bacterial methionyl-tRNA synthetase. Proceedings of the National Academy of Sciences of the United States of America. 2011;108(17):6933-8.

56. Ling J, Soll D. Severe oxidative stress induces protein mistranslation through impairment of an aminoacyl-tRNA synthetase editing site. Proceedings of the National Academy of Sciences of the United States of America. 2010;107(9):4028-33.

57. Santos MA, Cheesman C, Costa V, Moradas-Ferreira P, Tuite MF. Selective advantages created by codon ambiguity allowed for the evolution of an alternative genetic code in Candida spp. Molecular microbiology. 1999;31(3):937-47.

58. Bezerra AR, Simoes J, Lee W, Rung J, Weil T, Gut IG, et al. Reversion of a fungal genetic code alteration links proteome instability with genomic and phenotypic diversification. Proceedings of the National Academy of Sciences of the United States of America. 2013;110(27):11079-84.

59. Sherman F. Getting started with yeast. Methods in enzymology. 2002;350:3-41.

60. Khurana V, Lindquist S. Modelling neurodegeneration in Saccharomyces cerevisiae: why cook with baker's yeast? Nature reviews Neuroscience. 2010;11(6):436-49.

61. Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, et al. Life with 6000 genes. Science. 1996;274(5287):546, 63-7.

62. Karathia H, Vilaprinyo E, Sorribas A, Alves R. Saccharomyces cerevisiae as a model organism: a comparative study. PloS one. 2011;6(2):e16015.

63. Gietz RD, Woods RA. Genetic transformation of yeast. BioTechniques. 2001;30(4):816-20, 22-6, 28 passim.

64. Bloom-Ackermann Z, Navon S, Gingold H, Towers R, Pilpel Y, Dahan O. A comprehensive tRNA deletion library unravels the genetic architecture of the tRNA pool. PLoS genetics. 2014;10(1):e1004084.

65. Barelle CJ, Manson CL, MacCallum DM, Odds FC, Gow NA, Brown AJ. GFP as a quantitative reporter of gene regulation in Candida albicans. Yeast. 2004;21(4):333-40.

66. Sambrook J, Maniatis T, Fritsch EF, Laboratory CSH. Molecular cloning : a laboratory manual. 2nd ed: Cold Spring Harbor Laboratory Press; 1989.

67. Walther A, Wendland J. An improved transformation protocol for the human fungal pathogen Candida albicans. Current genetics. 2003;42(6):339-43.

68. Heitzler J, Marechal-Drouard L, Dirheimer G, Keith G. Use of a dot blot hybridization method for identification of pure tRNA species on different membranes. Biochimica et biophysica acta. 1992;1129(3):273-7.

69. Guo HH, Choe J, Loeb LA. Protein tolerance to random amino acid change. Proceedings of the National Academy of Sciences of the United States of America. 2004;101(25):9205-10.

70. Gomes AC, Miranda I, Silva RM, Moura GR, Thomas B, Akoulitchev A, et al. A genetic code alteration generates a proteome of high diversity in the human pathogen Candida albicans. Genome biology. 2007;8(10):R206.

71. Hoesl MG, Budisa N. Recent advances in genetic code engineering in Escherichia coli. Current opinion in biotechnology. 2012;23(5):751-7.

72. Cropp TA, Schultz PG. An expanding genetic code. Trends in genetics : TIG. 2004;20(12):625-30.

73. Xie J, Schultz PG. An expanding genetic code. Methods. 2005;36(3):227-38.

74. Chatterjee A, Xiao H, Schultz PG. Evolution of multiple, mutually orthogonal prolyl-tRNA synthetase/tRNA pairs for unnatural amino acid mutagenesis in Escherichia coli. Proceedings of the National Academy of Sciences of the United States of America. 2012;109(37):14841-6.

75. Wang L, Brock A, Herberich B, Schultz PG. Expanding the genetic code of Escherichia coli. Science. 2001;292(5516):498-500.

76. Anderson JC, Wu N, Santoro SW, Lakshman V, King DS, Schultz PG. An expanded genetic code with a functional quadruplet codon. Proceedings of the National Academy of Sciences of the United States of America. 2004;101(20):7566-71.

77. Pezo V, Metzgar D, Hendrickson TL, Waas WF, Hazebrouck S, Doring V, et al. Artificially ambiguous genetic code confers growth yield advantage. Proceedings of the National Academy of Sciences of the United States of America. 2004;101(23):8593-7.

78. Cowie DB, Cohen GN. Biosynthesis by Escherichia coli of active altered proteins containing selenium instead of sulfur. Biochimica et biophysica acta. 1957;26(2):252-61.

79. Yoshikawa E, Fournier MJ, Mason TL, Tirrell DA. Genetically Engineered Fluoropolymers. Synthesis of Repetitive Polypeptides Containing p-Fluorophenylalanine residues. Macromolecules. 1994;27(19):5471-5.

80. Budisa N, Steipe B, Demange P, Eckerskorn C, Kellermann J, Huber R. High-level biosynthetic substitution of methionine in proteins by its analogs 2-aminohexanoic acid, selenomethionine, telluromethionine and ethionine in Escherichia coli. European journal of biochemistry / FEBS. 1995;230(2):788-96.

81. Geslain R, Cubells L, Bori-Sanz T, Alvarez-Medina R, Rossell D, Marti E, et al. Chimeric tRNAs as tools to induce proteome damage and identify components of stress responses. Nucleic acids research. 2010;38(5):e30.

82. Henikoff S, Henikoff JG. Amino acid substitution matrices from protein blocks. Proceedings of the National Academy of Sciences of the United States of America. 1992;89(22):10915-9.

83. Woese CR, Olsen GJ, Ibba M, Soll D. Aminoacyl-tRNA synthetases, the genetic code, and the evolutionary process. Microbiology and molecular biology reviews : MMBR. 2000;64(1):202-36.

84. Lenhard B, Orellana O, Ibba M, Weygand-Durasevic I. tRNA recognition and evolution of determinants in seryl-tRNA synthesis. Nucleic acids research. 1999;27(3):721-9.

85. Weissenbach J, Dirheimer G, Falcoff R, Sanceau J, Falcoff E. Yeast tRNALeu (anticodon U--A--G) translates all six leucine codons in extracts from interferon treated cells. FEBS letters. 1977;82(1):71-6.

86. Randerath E, Gupta RC, Chia LL, Chang SH, Randerath K. Yeast tRNA Leu UAG. Purification, properties and determination of the nucleotide sequence by radioactive derivative methods. European journal of biochemistry / FEBS. 1979;93(1):79-94.

87. Rogalski M, Karcher D, Bock R. Superwobbling facilitates translation with reduced tRNA sets. Nature structural & molecular biology. 2008;15(2):192-8.

88. Johansson MJ, Esberg A, Huang B, Bjork GR, Bystrom AS. Eukaryotic wobble uridine modifications promote a functionally redundant decoding system. Molecular and cellular biology. 2008;28(10):3301-12.

89. Vimaladithan A, Farabaugh PJ. Special peptidyl-tRNA molecules can promote translational frameshifting without slippage. Molecular and cellular biology. 1994;14(12):8107-16.

90. Kawakami K, Pande S, Faiola B, Moore DP, Boeke JD, Farabaugh PJ, et al. A rare tRNA-Arg(CCU) that regulates Ty1 element ribosomal frameshifting is essential for Ty1 retrotransposition in Saccharomyces cerevisiae. Genetics. 1993;135(2):309-20.

91. Huang Q, Yao P, Eriani G, Wang ED. In vivo identification of essential nucleotides in tRNALeu to its functions by using a constructed yeast tRNALeu knockout strain. Nucleic acids research. 2012;40(20):10463-77.

92. New AM, Cerulus B, Govers SK, Perez-Samper G, Zhu B, Boogmans S, et al. Different levels of catabolite repression optimize growth in stable and variable environments. PLoS biology. 2014;12(1):e1001764.

93. Ma P, Goncalves T, Maretzek A, Dias MC, Thevelein JM. The lag phase rather than the exponential-growth phase on glucose is associated with a higher cAMP level in wild-type and cAPK-attenuated strains of the yeast Saccharomyces cerevisiae. Microbiology. 1997;143 (Pt 11):3451-9.

94. Rose M, Grisafi P, Botstein D. Structure and function of the yeast URA3 gene: expression in Escherichia coli. Gene. 1984;29(1-2):113-24.

95. Moura G, Pinheiro M, Silva R, Miranda I, Afreixo V, Dias G, et al. Comparative context analysis of codon pairs on an ORFeome scale. Genome biology. 2005;6(3):R28.

96. Caponigro G, Muhlrad D, Parker R. A small segment of the MAT alpha 1 transcript promotes mRNA decay in Saccharomyces cerevisiae: a stimulatory role for rare codons. Molecular and cellular biology. 1993;13(9):5141-8.

97. Pescini D, Cazzaniga P, Besozzi D, Mauri G, Amigoni L, Colombo S, et al. Simulation of the Ras/cAMP/PKA pathway in budding yeast highlights the establishment of stable oscillatory states. Biotechnology advances. 2012;30(1):99-107.

98. Tanaka K, Nakafuku M, Satoh T, Marshall MS, Gibbs JB, Matsumoto K, et al. S. cerevisiae genes IRA1 and IRA2 encode proteins that may be functionally equivalent to mammalian ras GTPase activating protein. Cell. 1990;60(5):803-7.

99. Haney SA, Broach JR. Cdc25p, the guanine nucleotide exchange factor for the Ras proteins of Saccharomyces cerevisiae, promotes exchange by stabilizing Ras in a nucleotide-free state. The Journal of biological chemistry. 1994;269(24):16541-8.

100. Thevelein JM, de Winde JH. Novel sensing mechanisms and targets for the cAMP-protein kinase A pathway in the yeast Saccharomyces cerevisiae. Molecular microbiology. 1999;33(5):904-18.

101. Yun CW, Tamaki H, Nakayama R, Yamamoto K, Kumagai H. Gpr1p, a putative G-protein coupled receptor, regulates glucose-dependent cellular cAMP level in yeast Saccharomyces cerevisiae. Biochemical and biophysical research communications. 1998;252(1):29-33.

102. Tyedmers J, Madariaga ML, Lindquist S. Prion switching in response to environmental stress. PLoS biology. 2008;6(11):e294.

103. Woese CR. On the evolution of the genetic code. Proceedings of the National Academy of Sciences of the United States of America. 1965;54(6):1546-52.

104. Barrick JE, Lenski RE. Genome dynamics during experimental evolution. Nature reviews Genetics. 2013;14(12):827-39.

105. Oxman E, Alon U, Dekel E. Defined order of evolutionary adaptations: experimental evidence. Evolution; international journal of organic evolution. 2008;62(7):1547-54.

106. Yona AH, Bloom-Ackermann Z, Frumkin I, Hanson-Smith V, Charpak-Amikam Y, Feng Q, et al. tRNA genes rapidly change in evolution to meet novel translational demands. eLife. 2013;2:e01339.

107. Yap LP, Stehlin C, Musier-Forsyth K. Use of semi-synthetic transfer RNAs to probe molecular recognition by Escherichia coli proline-tRNA synthetase. Chemistry & biology. 1995;2(10):661-6.

108. Singh R, Green MR. Sequence-specific binding of transfer RNA by glyceraldehyde-3-phosphate dehydrogenase. Science. 1993;259(5093):365-8.

109. Becker HF, Motorin Y, Sissler M, Florentz C, Grosjean H. Major identity determinants for enzymatic formation of ribothymidine and pseudouridine in the T psi-loop of yeast tRNAs. Journal of molecular biology. 1997;274(4):505-18.

110. Grosjean H, Sprinzl M, Steinberg S. Posttranscriptionally modified nucleosides in transfer RNA: their locations and frequencies. Biochimie. 1995;77(1-2):139-41.

111. Lecointe F, Namy O, Hatin I, Simos G, Rousset JP, Grosjean H. Lack of pseudouridine 38/39 in the anticodon arm of yeast cytoplasmic tRNA decreases in vivo recoding efficiency. The Journal of biological chemistry. 2002;277(34):30445-53.

112. Charette M, Gray MW. Pseudouridine in RNA: what, where, how, and why. IUBMB life. 2000;49(5):341-51.

113. Becker HF, Motorin Y, Planta RJ, Grosjean H. The yeast gene YNL292w encodes a pseudouridine synthase (Pus4) catalyzing the formation of psi55 in both mitochondrial and cytoplasmic tRNAs. Nucleic acids research. 1997;25(22):4493-9.

114. Cormack BP, Bertram G, Egerton M, Gow NA, Falkow S, Brown AJ. Yeast-enhanced green fluorescent protein (yEGFP): a reporter of gene expression in Candida albicans. Microbiology. 1997;143 (Pt 2):303-11.

115. Wood TI, Barondeau DP, Hitomi C, Kassmann CJ, Tainer JA, Getzoff ED. Defining the role of arginine 96 in green fluorescent protein fluorophore biosynthesis. Biochemistry. 2005;44(49):16211-20.

116. Boeke JD, LaCroute F, Fink GR. A positive selection for mutants lacking orotidine-5'-phosphate decarboxylase activity in yeast: 5-fluoro-orotic acid resistance. Molecular & general genetics : MGG. 1984;197(2):345-6.

117. Sheltzer JM, Blank HM, Pfau SJ, Tange Y, George BM, Humpton TJ, et al. Aneuploidy drives genomic instability in yeast. Science. 2011;333(6045):1026-30.

118. Pavelka N, Rancati G, Zhu J, Bradford WD, Saraf A, Florens L, et al. Aneuploidy confers quantitative proteome changes and phenotypic variation in budding yeast. Nature. 2010;468(7321):321-5.

119. Dataset: Codon usage tables [27.10.2014]. Tables for S. cerevisiae are based on SGD genome version R64-1-1 (January 2011).]. Available from: http://downloads.yeastgenome.org/unpublished_data/codon/64_1_1_all_nuclear.cusp.

120. Matsumoto K, Uno I, Oshima Y, Ishikawa T. Isolation and characterization of yeast mutants deficient in adenylate cyclase and cAMP-dependent protein kinase. Proceedings of the National Academy of Sciences of the United States of America. 1982;79(7):2355-9.

121. Robinson LC, Gibbs JB, Marshall MS, Sigal IS, Tatchell K. CDC25: a component of the RAS-adenylate cyclase pathway in Saccharomyces cerevisiae. Science. 1987;235(4793):1218-21.

122. Pavon-Eternod M, Gomes S, Geslain R, Dai Q, Rosner MR, Pan T. tRNA over-expression in breast cancer and functional consequences. Nucleic acids research. 2009;37(21):7268-80.

123. Winter AG, Sourvinos G, Allison SJ, Tosh K, Scott PH, Spandidos DA, et al. RNA polymerase III transcription factor TFIIIC2 is overexpressed in ovarian tumors. Proceedings of the National Academy of Sciences of the United States of America. 2000;97(23):12619-24.

124. Daly NL, Arvanitis DA, Fairley JA, Gomez-Roman N, Morton JP, Graham SV, et al. Deregulation of RNA polymerase III transcription in cervical epithelium in response to high-risk human papillomavirus. Oncogene. 2005;24(5):880-8.

125. Pavon-Eternod M, Gomes S, Rosner MR, Pan T. Overexpression of initiator methionine tRNA leads to global reprogramming of tRNA expression and increased proliferation in human epithelial cells. Rna. 2013;19(4):461-6.

126. Wahab SZ, Rowley KO, Holmes WM. Effects of tRNA(1Leu) overproduction in Escherichia coli. Molecular microbiology. 1993;7(2):253-63.

127. Lin JP, Aker M, Sitney KC, Mortimer RK. First position wobble in codon-anticodon pairing: amber suppression by a yeast glutamine tRNA. Gene. 1986;49(3):383-8.

128. Pure GA, Robinson GW, Naumovski L, Friedberg EC. Partial suppression of an ochre mutation in Saccharomyces cerevisiae by multicopy plasmids containing a normal yeast tRNAGln gene. Journal of molecular biology. 1985;183(1):31-42.

129. Thompson RC, Karim AM. The accuracy of protein biosynthesis is limited by its speed: high fidelity selection by ribosomes of aminoacyl-tRNA ternary complexes containing GTP[gamma S]. Proceedings of the National Academy of Sciences of the United States of America. 1982;79(16):4922-6.

130. Wohlgemuth I, Pohl C, Mittelstaet J, Konevega AL, Rodnina MV. Evolutionary optimization of speed and accuracy of decoding on the ribosome. Philosophical transactions of the Royal Society of London Series B, Biological sciences. 2011;366(1580):2979-86.

131. Silva RM, Paredes JA, Moura GR, Manadas B, Lima-Costa T, Rocha R, et al. Critical roles for a genetic code alteration in the evolution of the genus Candida. The EMBO journal. 2007;26(21):4555-65.

Annexs

Drop-out mix:

0.25g Adenine
1 g Alanine
1 g Arginine
1 g Asparagine
1 g Aspartate
1 g Cysteine
1 g Glycine
1 g Histidine
1 g Inositol
1 g Isoleucine
1 g Glutamate
1 g Glutamine
1 g Lysine
1 g Leucine
1 g Methionine
1 g Phenylalanine
1 g Proline
1 g Serine
1 g Threonine
1 g Tryptophan
1 g Tyrosine
1 g Uracil
1 g Valine

5-FOA Plates (1 L):

20 g Agar
500 ml dH_2O
Autoclaved separately
6.7 g Bacto-yeast Nitrogen Base
2 g Drop-out –Ura
20 g Glucose
50 mg Uracil (50 µg/ml)
1 g 5-FOA
500 ml dH_2O
Filtered by 0.02 µm filters and added to the agar when cooled to ~ 55ºC

Plasmid Maps

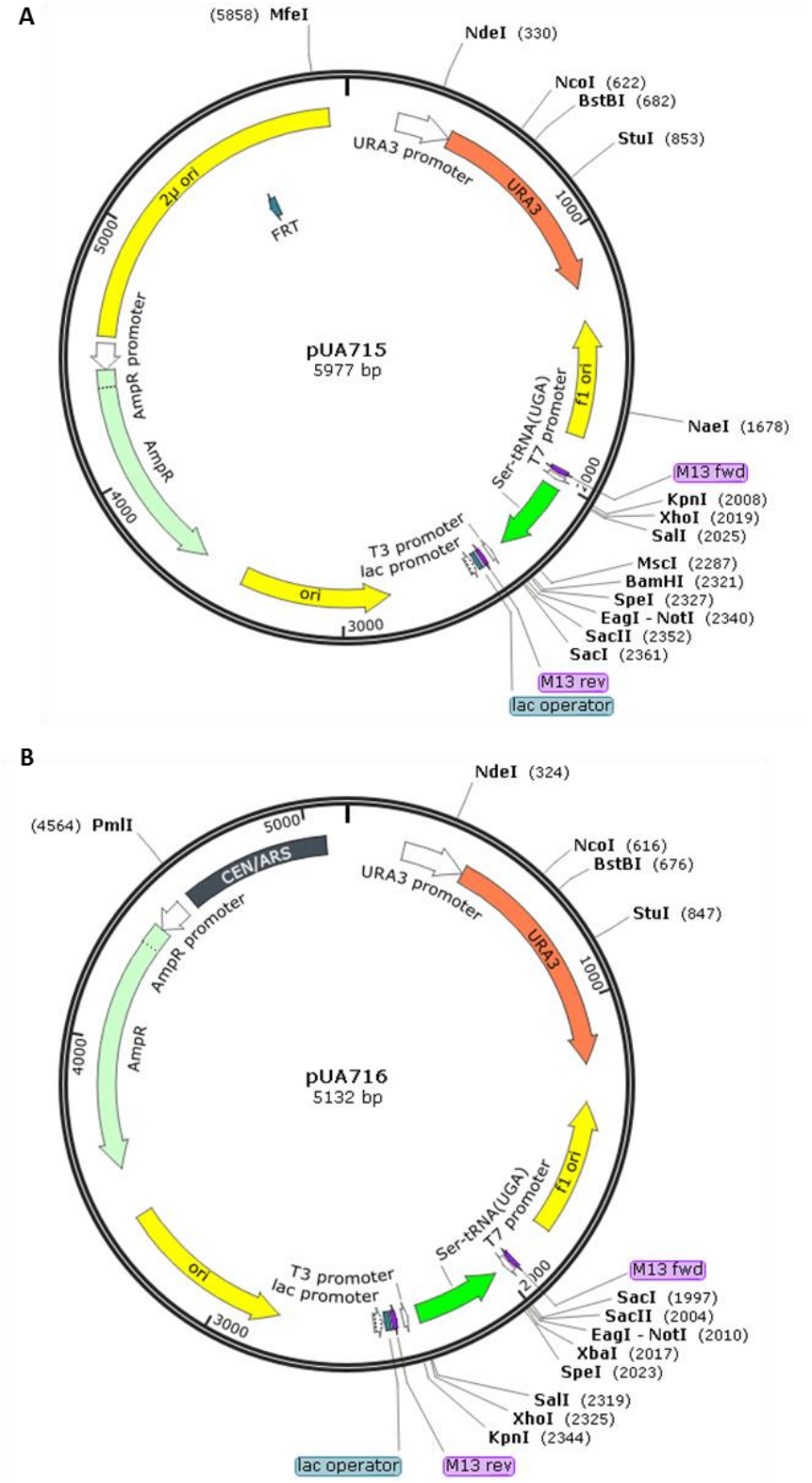


Figure 2.1 – Maps of the constructed plasmids pUA715 and pUA716 used as backbone for expression of mutant tRNAs, with insertion of the *C. albicans* $tRNA^{Ser}_{UGA}$ gene between *SalI* and *BamHI* in the multi-copy vector pRS426 (A), and between *SpeI* and *SalI* in the single-copy vector pRS316 (B).

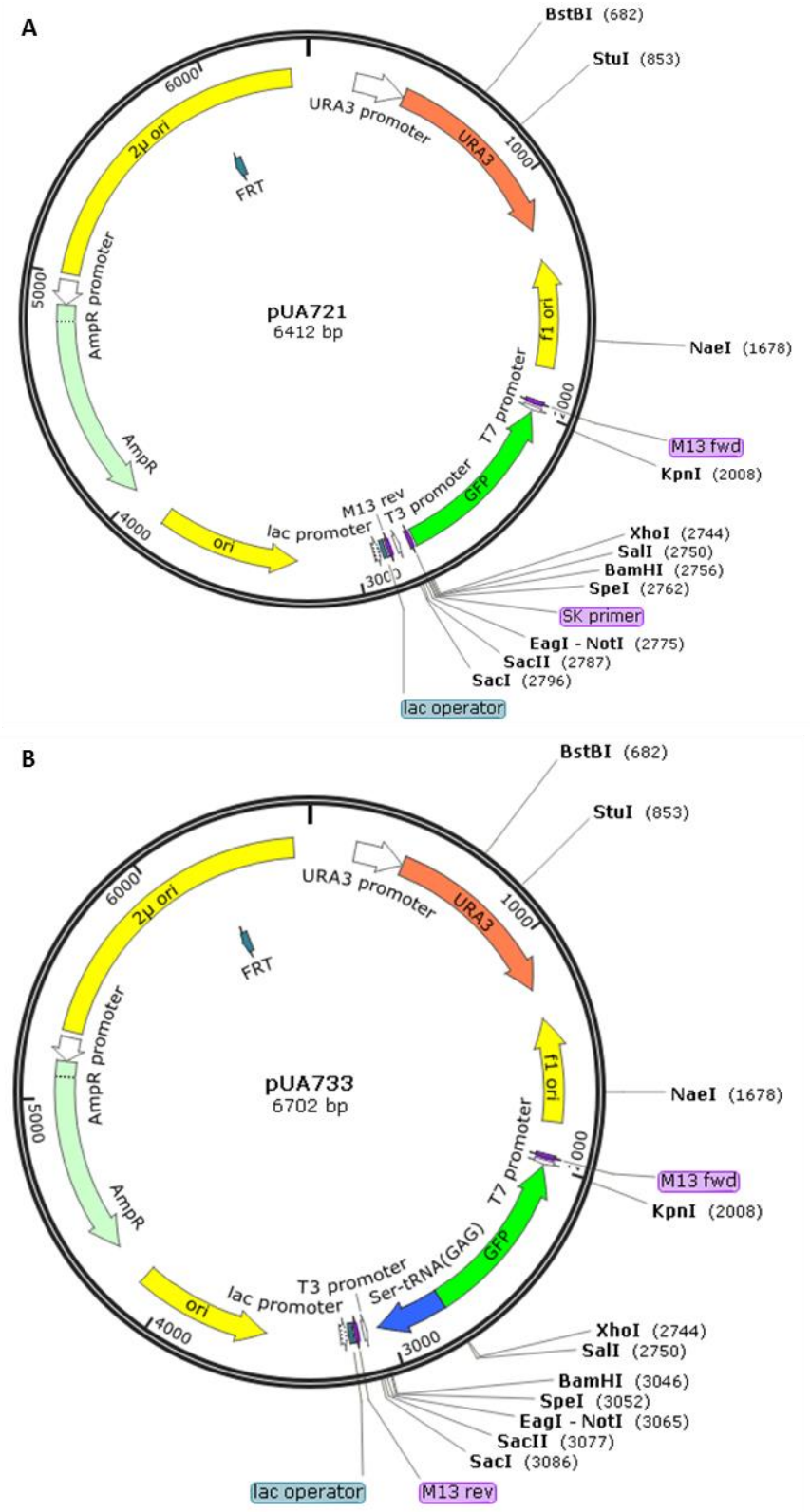


Figure 2.2 – Map of the constructed reporters pUA721 and pUA733. (A) Backbone for control plasmids, only with GFP inserted between *KpnI* and *XhoI* (positions of interest were posteriorly mutated).(B) Backbone for the reporter construction, with insertion of the mutagenized $tRNA^{Ser}_{GAG}$ gene between the restriction sites *SalI* and *BamHI*, and the *y*E-GFP gene between *KpnI* and *XhoI* (positions of interest were posteriorly mutated).

Table 3.3 – Complete list of PKA related genes and CUC and AGG codon frequency

Gene	Description	No. CUC codons	No. AGG codons
IRA2	GTPase-activating protein	22	31
IRA1	GTPase-activating protein	19	30
CYR1	Adenylate cyclase	16	25
RIM15	Protein kinase involved in cell proliferation	15	21
TOR1	Phosphatidylinositol kinase-related protein kinase	13	23
CCC2	Cu^{2+}-transporting P-type ATPase	10	6
GEX2	Proton:glutathione antiporter	10	4
GPR1	Plasma membrane G-protein coupled receptor	10	5
TOR2	Phosphatidylinositol kinase-related protein kinase	10	37
GEX1	Proton:glutathione antiporter	9	4
CDC25	Ras-guanine exchange factor	6	20
ROM2	GDP/GTP exchange factor for Rho proteins	6	16
BCY1	cAMP-dependent protein kinase regulatory subunit	5	5
GPB1	Multistep regulator of cAMP-PKA signaling	5	13
GPB2	Multistep regulator of cAMP-PKA signaling	5	7
MSN2	Transcriptional activator, activated in stress conditions	5	6
ATG13	Regulatory subunit of the Atg1p signaling complex	4	5
GPA2	Nucleotide binding alpha subunit of the heterotrimeric G protein Gpr1p	4	8
KSP1	Serine/threonine protein kinase	4	13
PHO80	Cyclin, interacts with cyclin-dependent kinase Pho85p	4	1
SCH9	AGC (PKA, PKC, PKG) family protein kinase	4	3
UBP3	Ubiquitin-specific protease involved in transport and osmotic response	4	4
YAK1	Serine-threonine protein kinase	4	11
ATG1	Protein serine/threonine kinase	3	11
PDE2	3',5'-cyclic-nucleotide phosphodiesterase	3	3
PHO85	cyclin-dependent serine/threonine-protein kinase	3	2
WHI3	RNA binding protein that sequesters CLN3 mRNA	3	4
YPL260W	Putative substrate of cAMP-dependent protein kinase	3	5
COM2	Transcription factor that binds IME1 Upstream Activation Signal	2	9
FLO11	GPI-anchored cell surface glycoprotein (flocculin)	2	0

LYS1	saccharopine dehydrogenase (NAD^+, L-lysine-forming)	2	1
PAT1	Deadenylation-dependent mRNA-decapping factor	2	5
PDE1	3',5'-cyclic-nucleotide phosphodiesterase	2	3
SPI1	GPI-anchored cell wall protein	2	0
YPK3	AGC kinase	2	5
GAS1	1,3-beta-glucanosyl transferase	1	0
MYO2	Type V myosin motor involved in actin-based transport	1	12
RAS1	GTPase involved in G-protein signaling in adenylate cyclase activation	1	2
SLT2	Mitogen-activated serine/threonine-protein kinase	1	6
SNF1	AMP-activated serine/threonine-protein kinase catalytic subunit	1	5
SOK2	Nuclear protein that negatively regulates pseudohyphal differentiation	1	5
TOM22	Component of the TOM (translocase of outer membrane) complex	1	0
TPK2	Ras family GTPase	1	5
CDC19	Pyruvate kinase	0	0
CYC1	Cytochrome-c isoform 1	0	0
HSP12	Lipid-binding protein	0	0
PHO84	High-affinity inorganic phosphate (Pi) transporter	0	0
RAS2	Ras family GTPase	0	5
TOM40	Component of the TOM (translocase of outer membrane) complex	0	0
TPK1	cAMP-dependent protein kinase catalytic subunit Tpk1	0	5
TPK3	cAMP-dependent protein kinase catalytic subunit Tpk3	0	3
WHI2	Protein required for full activation of the general stress response	0	4